Investigations into the effect of Hadron Collider data on MSTW Parton Distribution Functions

Benjamin Joseph Alexander Watt
University College London

Submitted to University College London in fulfilment
of the requirements for the award of the
degree of **Doctor of Philosophy**

October 11 2013

Declaration

I, Benjamin Joseph Alexander Watt confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Benjamin Watt _____

Abstract

The latest jet data from the Large Hadron Collider is studied in the context of the MSTW Parton Distribution Functions (PDFs). The effect of recent ATLAS inclusive- and dijet results on the current standard PDF sets is investigated, compared and contrasted to similar results from the Tevatron. A parton reweighting technique is used to gain information on the optimum PDFs for each new data set.

New PDF sets are produced and studied using the new LHC data. These jet results provide a new central value of the PDF and reduce the uncertainty on the distributions. Additionally, a new method of parametrising the quark PDFs using Chebyshev Polynomials is tested in relation to the ATLAS W/Z rapidity data, which is described poorly using the standard PDFs. The effect of parton showering on jet physics is studied using various shower Monte Carlo generators within the context of jets produced in deep inelastic scattering, and the possibility of using charged current jet production in PDF fits is tested.

Acknowledgements

There are many people without whom this PhD would not have been possible. I would like to thank my supervisor Robert Thorne for his immense help and guidance throughout the past 4 years. It was a privilege to work with you in these fascinating investigations into the proton. Also Patrick Motylinski, whose collaboration in the final year of my thesis was vital for the results presented. My time working in Germany under the MCnet program was a fantastic opportunity and wonderful experience, so my thanks go to Stefan Gieseke and the ITP in Karlsruhe for their help during that project.

The UCL HEP group in general has provided a superb working environment. Thanks go to my fellow students and the academics for making the group what it is. Friday nights at the JB will be sorely missed.

Finally, and most importantly, I want to thank my mum and dad. Their love and support has helped me to achieve not only this PhD, but everything before and whatever comes after.

Contents

List of Figures

List of Tables

Chapter 1.

Theoretical Framework

1.1. The Standard Model

The Standard Model of particle physics is the result of decades of theoretical and experimental work into the fundamental interactions of matter. It successfully describes the properties of quarks, leptons and the force carrying bosons which have been observed in nature: the photon, $W^{\pm}/Z$ and gluon. However, the observation of neutrino mass was the first hint that there is physics beyond the Standard Model, in which the neutrinos are required to be massless.

The Lagrangian of the Standard Model is constructed using three Lie groups, $SU(3), SU(2), U(1)$ which represent the strong and electroweak forces. Gravity is not included in the model, and the inclusion of this force remains an unresolved issue in Quantum Field Theory. The Lie groups describe the underlying symmetry under gauge transformations of the forces, and thus their properties.

The Lagrangian of the Standard Model can be written as:

$$\mathcal{L} = \mathcal{L}_{QCD} + \mathcal{L}_{EW} + \mathcal{L}_{H} \tag{1.1}$$

where

$$\mathcal{L}_{EW} = \sum_{\psi} \bar{\psi}\gamma^{\mu}(i\partial_{\mu} - g'\frac{1}{2}Y_W B_{\mu} - g\frac{1}{2}\tau W_{\mu}) - \frac{1}{4}F^{a}_{\mu\nu}F^{\mu\nu}_{a} \tag{1.2}$$

$$\mathcal{L}_{QCD} = \bar{\psi}_i(i\not{D}_{ij} - m\delta_{ij})\psi_j - \frac{1}{4}G^{a}_{\mu\nu}G^{\mu\nu}_{a} \tag{1.3}$$

$$\mathcal{L}_{H} = \phi^{\dagger}(\partial^{\mu} - \frac{i}{2}(g'Y_W B^{\mu} + g\tau W^{\mu}))(\partial_{\mu} + \frac{i}{2}(g'Y_W B_{\mu} + g\tau W_{\mu}))\phi - \frac{\lambda^2}{4}(\phi^{\dagger}\phi - v^2)^2 \tag{1.4}$$

Until July 2012, the Higgs sector was not experimentally confirmed, but upon discovery of a boson at both the ATLAS and CMS experiments with Higgs-like properties, it is all but confirmed as the source of electroweak symmetry breaking.

The particles of the Standard Model are excitations of the fields included in the Lagrangian. The QCD sector describes a group of fermions called quarks, which possess both colour and electromagnetic charge, and the massless bosons which carry the colour charge, namely gluons. The electromagnetic sector includes the fermions called leptons and the force carrier, the photon. A summary of the particles of the Standard Model is represented in Table 1.1.

				Bosons	
Quarks	u	c	t	γ	H
	d	s	b	g	
Leptons	ν_e	ν_μ	ν_τ	Z_0	
	e	μ	τ	$W^{\pm}$	

Table 1.1.: The particles of the Standard Model.

1.2. Quantum Chromodynamics

The section of the Standard Model which is of most importance to this thesis is that describing the strong nuclear interaction, known as Quantum Chromodynamics (QCD). The fundamental fields involved in the interaction are the 3 generations of quarks and a vector boson named the gluon.

When considering the interactions of quarks and gluons, such as in QCD, one needs to use a gauge theory with non-commuting gauge parameters. This is a very different theory to the Abelian QED (Quantum Electrodynamics) case, where the gauge fields commute and the gauge bosons consequently are not self-interacting. The gauge boson fields in QCD are $A^a_\mu(x)$, where a is the SU(n) gauge index of the theory. In order to construct the theory, the fundamental quantity required is the Lie algebra, which defines the commutation relation of the infinitesimal gauge generators. Namely

$$[T^a, T^b] = if^{abc}T^c \tag{1.5}$$

The generators T^a are defined by the gauge transformation

$$A_\mu(x) \rightarrow A'_\mu(x) = U^{-1}(x)A_\mu(x)U(x) + \frac{i}{g}U^{-1}(x)\partial_\mu U(x) \tag{1.6}$$

where

$$U(x) = \exp(-i\alpha^a(x)T^a) \tag{1.7}$$

This is in fact much more complicated that the Abelian case due to the fact that the generators and the gauge fields are Hermitian matrices, and so do not commute with each other. Hence, the infinitesimal form of the transformation contains an extra term dependent on the gauge fields

$$\begin{aligned} A^a_\mu(x) \to A'^a_\mu(x) &= A^a_\mu(x) + \frac{1}{g}\partial_\mu\alpha^a(x) - igA^b_\mu(T^b)_{ac}\alpha^c \\ &= A^a_\mu(x) + \frac{1}{g}\partial_\mu\alpha^a(x) + gf^{abc}A^b_\mu\alpha^c \\ &= A^a_\mu(x) + \frac{1}{g}(D_\mu\alpha(x))^a \end{aligned} \tag{1.8}$$

where the covariant derivative has been defined in the adjoint representation as

$$(D_\mu\alpha)^a = \partial_\mu\alpha^a(x) + gf^{abc}A^b_\mu\alpha^c. \tag{1.9}$$

There is more than one representation for this Lie algebra, and so a choice must be made. The fundamental representation is a set of $n \times n$ matrices which satisfy the algebra, and are the most basic objects which can be used. However, another useful representation is the adjoint representation. The form of the generators in this case are given by the relation

$$(T_a)_{bc} = -if_{abc} \tag{1.10}$$

In explicit calculations, knowledge of the explicit values of these objects is rarely needed. The more important quantities are the traces over colour indices of combinations of the generators. These are

$$C(R) = tr(T^aT^b) \quad , \quad C_2(R) = tr(T^aT^a) \tag{1.11}$$

where R is the representation. These two quantities can be thought of as normalisation constants which need to be set. The standard convention is to set, in the fundamental representation for SU(n),

$$C(F) = \frac{1}{2} \quad , \quad C_2(F) = \frac{n^2 - 1}{2n}. \tag{1.12}$$

And hence, in the adjoint representation

$$C(A) = n = C_2(A). \tag{1.13}$$

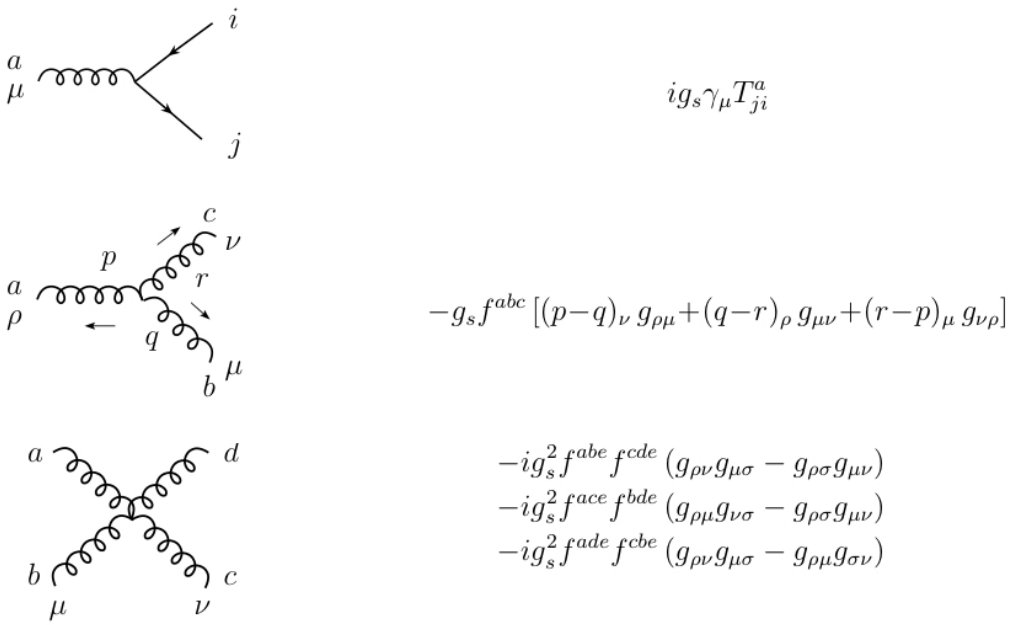


Figure 1.1.: Feynman rules for QCD.

The Feynman rules of interactions in QCD are shown in Fig 1.1 in the Feynman gauge, which is defined by the gauge fixing condition $\partial_\mu A^\mu = 0$. The presence of the 3- and 4-point self interaction vertices of the gluon are a result of the non-Abelian property of QCD, and will be important in the following discussion of PDFs.

Quark	Charge	Mass	Isospin
u	$+\frac{2}{3}$	$\sim$ 2 MeV	$+\frac{1}{2}$
d	$-\frac{1}{3}$	~ 5 MeV	$-\frac{1}{2}$
c	$+\frac{2}{3}$	~ 1.3 GeV	0
s	$-\frac{1}{3}$	~ 100 MeV	0
t	$+\frac{2}{3}$	~ 173 GeV	0
b	$-\frac{1}{3}$	~ 4 GeV	0

Table 1.2.: The 6 quarks involved in the strong force.

There are three generations of quarks which take part in the strong interaction, each with a unique set of properties. The quark charges are all either $+2/3$ or $-1/3$, allowing for hadrons with integer values of charge, either by combining 3 quarks into a baryon or a quark and an antiquark into a meson. The properties of each quark generation are shown in Table 1.2. The large differences in mass lead to a two categories of quark: light and heavy, the latter of which generally being defined as having mass greater than Λ_{QCD}, the scale below which perturbation theory is no longer valid.

1.3. Parton Distribution Functions

For any interaction involving hadrons in the initial state, such as those at HERA, the Tevatron and the LHC, one must have a reliable theory to describe which partons take part in the collision. In order to do this, the binding of quarks and gluons at low energies must be understood. Unfortunately, it is impossible to use perturbation theory to describe these low-energy interactions, due to the nature of the coupling constant in QCD, α_s. As described in the previous section, the generators of the QCD Lagrangian do not commute, unlike those for QED. The mathematical consequence of this property leads to a self-interaction of the gluon with itself, something which does not occur for the photon. Thus, there are new diagrams to consider for any vertex including quarks and gluons, and so the running coupling within the renormalisation of QCD has a very different behaviour to that of QED. The 1-loop coupling in QCD can be written at an arbitrary scale as

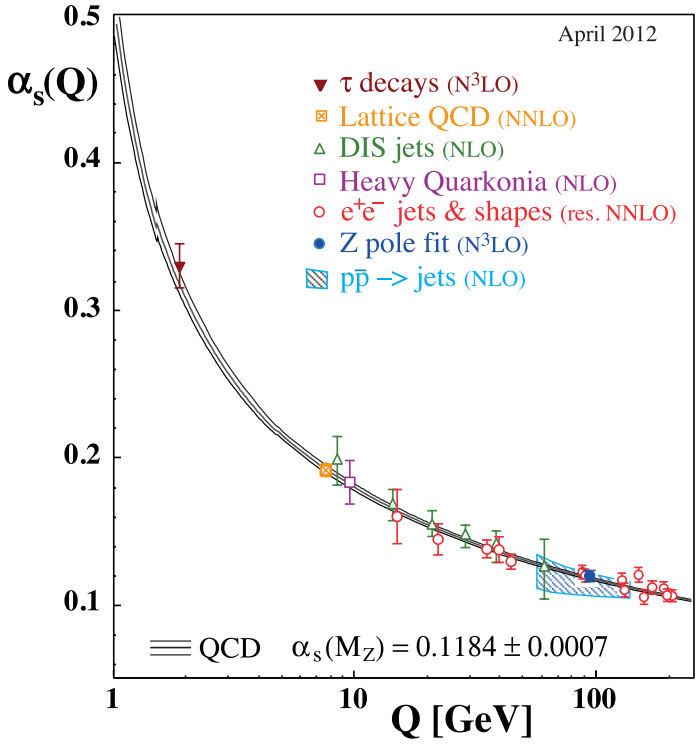


Figure 1.2.: The running of α_s, taken from [1].

$$\alpha_s(\mu_R^2) = \frac{\alpha_s(\mu_0^2)}{1 + \alpha_s(\mu_0^2)\beta_0 \ln \frac{\mu_R^2}{\mu_0^2}} \tag{1.14}$$

where the input scale μ_0 is a constant, and $\beta_0 = (33 - 2n_f)(12\pi)$ is the 1-loop beta function, which is positive for fewer than 16 flavours, n_f, of quarks, and hence for the observed 6 flavours of quarks. This functional form demonstrates the two phenomena known as asymptotic freedom and confinement in QCD. The coupling $\alpha_s \to 0$ as $\mu_R \to \infty$, whereas $\alpha_s \to \infty$ as $\mu_R \to \Lambda_{QCD}$.

Fig 1.2 demonstrates various experimentally measured values of α_s at different scales, and the logarithmic property can be seen clearly. The consequence of this behaviour leads to the breakdown of perturbation theory for processes with an energy transfer below $\Lambda_{QCD} \sim 200$MeV. Thus, the binding of partons inside a proton is inaccessible to Quantum Field Theory, and another method must be used to extract information on the struck parton in a proton collision. The only usable method currently is to simply use experimental data to determine the structure of the proton.

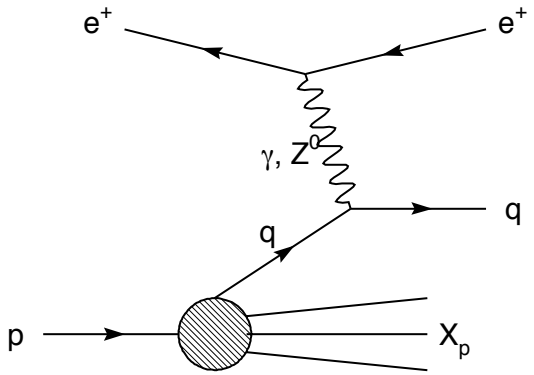


Figure 1.3.: Leading order DIS process.

In order to resolve this structure, one must use a small enough probe. The De Broglie wavelength, $\lambda = h/p$, determines the energy transfer required for a particle to probe a given length scale. An energy transfer of 100GeV will resolve a length scale of 0.02 fm, which is sufficiently below the proton radius to observe the parton structure.

An example of a process involving an initial state hadron is that of Deep Inelastic Scattering (DIS), first observed at SLAC-MIT [2]. This involves the collision of a lepton with a hadron, and such events were studied in *ep* interactions at the HERA collider in DESY. The electron interacts with a parton within the proton through the exchange of a highly virtual photon, $W^{\pm}$ or Z boson. In order to calculate such processes in Quantum Field Theory, the hadronic interaction is contained in a tensor $H_{\mu\nu}$. The diagram in Fig 1.3, which demonstrates a leading order interaction between an electron and a quark, can then be written as:

$$\frac{d\sigma}{dxdy} = \frac{2\pi y\alpha^2}{Q^4} L_{\mu\nu} H^{\mu\nu} \tag{1.15}$$

where Q^2 is the energy transfer in the process and $y = \frac{p \cdot q}{p \cdot k}$ is the inelasticity of the event. The leptonic tensor $L_{\mu\nu}$ can be calculated in QED, since it only contains interactions involving leptons. In addition to this, current conservation of the hadronic tensor ($q_\mu W^{\mu\nu} = 0$) can be applied to constrain the cross section to be proportional

to two unknown parameters W_1 and W_2. A third function, W_3 originates from the weak interaction involving the W and Z bosons, and is omitted for simplicity):

$$W^{\mu\nu} = W_1 \left(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2} \right) + \frac{W_2}{m^2} \left(p^\mu - \frac{p \cdot q}{q^2} q^\mu \right) \left(p^\nu - \frac{p \cdot q}{q^2} q^\nu \right) \tag{1.16}$$

The factors W_1 and W_2 are unknown functions of x and Q^2 which can only be found through experimental fits. Inputting this general expression into Eq.(1.15) along with the known leptonic tensor gives the DIS cross section as:

$$\frac{d\sigma}{dxdy} = \frac{2\pi\alpha^2}{xQ^4} \left[(1 + (1-y)^2 F_2(x, Q^2) - y^2 F_L(x, Q^2) \right] \tag{1.17}$$

where $F_2 = \frac{(p \cdot q) W_2}{m^2}$ and $F_L = F_2 - 2xF_1$ is the longitudinal structure function.

By applying the factorisation theorem, which states that the overall process can be decomposed into a parton-level subprocess convoluted with the appropriate parton distribution function, summed over all possible incoming partons inside the proton, the cross section can be written as

$$\frac{d\sigma}{dxdy} = \int d\xi f_a(\xi) \frac{d\hat{\sigma}}{dxdy} \tag{1.18}$$

where $\hat{\sigma}$ represents the purely partonic cross section, and ξ is the momentum fraction of the parent hadron carried by the struck parton. Comparing this form to the previously derived Eq.(1.17), the structure functions can be represented as a convolution of process-dependent coefficient functions and universal parton distribution functions (PDFs) $f_{q/H}$ of a parton q inside a hadron H.

$$F_i(x, Q^2) = \sum_{a=q,g} C_{i,a} \otimes f_{a/P}(Q^2) \tag{1.19}$$

Once calculated, the PDFs $f_q(Q^2)$ can be used in any other process involving a proton, and for processes involving two proton beams. The only change required in studying such processes is to calculate the relevant coefficient functions, and to convolute two separate PDFs. For example a generic cross section at the LHC would be calculated through

$$\sigma_{P_1P_2} = \sum_{a,b=q,g} C_{i,a,b}(\mu_R^2, \mu_F^2) \otimes f_{a/P_1}(\mu_F^2) \otimes f_{b/P_2}(\mu_F^2) \tag{1.20}$$

Once the PDFs have been determined at an input scale of e.g. 1 GeV, they can be evolved to any other scale by using the DGLAP evolution equations [3]:

$$\mu_F^2 \frac{\partial q_i(x,\mu_F^2)}{\partial \mu_F^2} = \int_x^1 \frac{d\xi}{\xi} P_{ij}\left(\frac{x}{\xi}, \alpha_s(\mu_R^2)\right) q_j(\xi, \mu_F^2) \tag{1.21}$$

The splitting functions $P_{ij}(x, Q^2)$ are the DGLAP evolution kernels which are motivated by the quantum field theoretical probability of a parton i emitting a parton j at the scale Q^2. The leading order splitting functions are shown in Table 1.3. Due to the presence of quark-gluon splitting functions, it is clear that the evolution of the gluon is intimately related to that of the quarks in a non-trivial manner.

1.3.1. MSTW method

In order to obtain the process independent PDFs, a global analysis of all hadron collision data must be performed. These data sets include fixed target and collider DIS, and hadron-hadron experiments. There are several groups currently producing PDFs, each differing in choice of input parametrisation, error propagation methods and selected data sets. The MSTW [4] set of PDFs is the subject of this thesis, of which the most recent official version was released in 2008. Other major PDF groups are NNPDF [5], HERAPDF [6],CT10 [7] and ABM [8].

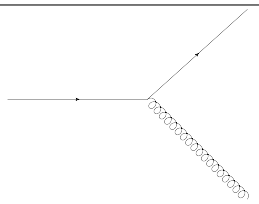
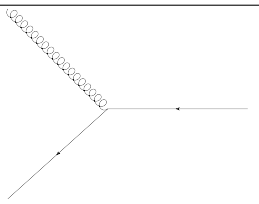
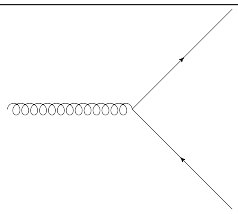
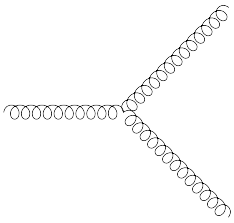
	$\frac{4}{3}\left[\frac{1+z^2}{1-z}\right]_+ + \frac{3}{2}\delta(1-x)$
	$\frac{4}{3}\frac{1+(1-z)^2}{z}$
	$\frac{1}{2}[z^2+(1-z)^2]$
	$6\left[\frac{1-z}{z} + \frac{z}{(1-z)_+} + z(1-z)\right] + \frac{33-n_f}{6}\delta(1-x)$

Table 1.3.: QCD Splitting Functions at leading order. The plus distribution in the first and last functions is defined as $\int_0^1 \frac{f(x)}{(1-x)_+}dx = \int_0^1 \frac{f(x)-f(1)}{1-x}dx$.

The parametrisation of the MSTW PDFs is shown in Eq.(1.22). In total there are 34 parameters including α_s. However, the sum rules in Eq.(1.23) constrain A_g, A_u, A_d and x_0. In addition, δ_- is set at 0.2 due to correlation between it and A_-. Finally, $\delta_S = \delta_s$ is set, leaving 28 parameters left free in the fit.

$$
\begin{aligned}
xu_v(x,Q_0^2) &= A_u x^{\eta_1}(1-x)^{\eta_2}(1+\epsilon_u\sqrt{x}+\gamma_u x)\\
xd_v(x,Q_0^2) &= A_d x^{\eta_3}(1-x)^{\eta_4}(1+\epsilon_d\sqrt{x}+\gamma_d x)\\
xS(x,Q_0^2) &= A_S x^{\delta_s}(1-x)^{\eta_S}(1+\epsilon_S\sqrt{x}+\gamma_S x)\\
x\Delta(x,Q_0^2) &= A_\Delta x^{\eta_\Delta}(1-x)^{\eta_{S+2}}(1+\gamma_\Delta x+\delta_\Delta x^2)\\
xg(x,Q_0^2) &= A_g x^{\delta_g}(1-x)^{\eta_g}(1+\epsilon_g\sqrt{x}+\gamma_g x)+A_{g'}x^{\delta_{g'}}(1-x)^{\eta_{g'}}\\
x(s+\bar{s})(x,Q_0^2) &= A_+ x^{\delta_S}(1-x)^{\eta_+}(1+\epsilon_S\sqrt{x}+\gamma_S x)\\
x(s-\bar{s})(x,Q_0^2) &= A_- x^{\delta_-}(1-x)^{\eta_-}(1-\frac{x}{x_0})
\end{aligned}
\tag{1.22}
$$

$$
\int_0^1 dx u_v(x,Q_0^2)=2,\ \int_0^1 dx d_v(x,Q_0^2)=1,\ \int_0^1 dx s_v(x,Q_0^2)=0 \tag{1.23}
$$

$$
\int_0^1 dx x[u_v(x,Q_0^2)+d_v(x,Q_0^2)+S(x,Q_0^2)+g(x,Q_0^2)]=1 \tag{1.24}
$$

The form of the parametrisation is motivated by some physics concepts. Each parton includes a term proportional to x, with the exception of the gluon which contains two proportional to x, and one proportional to $1-x$, such that the probabilities approach zero for partons carrying the entire proton momentum, and obey single power Regge theory for low momentum fraction.

The processes which are included in the fits are shown in Table 1.4, along with the partons probed and the range of x which is sensitive. The lowest x values are only probed by HERA data.

For each process included, the coefficient functions $C_{i,a}$ must be calculated as a function of the incoming parton momentum, such that the convolution into the full cross section and consequent PDF analysis can be performed.

The MSTW fitting code employs an iterative minimisation algorithm in order to obtain the optimum values of the parameters. At each stage, a PDF is produced using the current parameters, and the total χ^2 for all points in the fit is calculated for this intermediate PDF. In order to do this, it must be possible to calculate each theory cross section quickly for an arbitrary PDF and scale choice. For most cross

Process	Subprocess	Partons	x range
$\ell^{\pm}\{p,n\} \to \ell^{\pm} X$	$\gamma^* q \to q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
$\ell^{\pm} n/p \to \ell^{\pm} X$	$\gamma^* d/u \to d/u$	d/u	$x \gtrsim 0.01$
$pp \to \mu^+\mu^- X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
$pn/pp \to \mu^+\mu^- X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
$\nu(\bar{\nu})\, N \to \mu^-(\mu^+)\, X$	$W^* q \to q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
$\nu\, N \to \mu^-\mu^+ X$	$W^* s \to c$	s	$0.01 \lesssim x \lesssim 0.2$
$\bar{\nu}\, N \to \mu^+\mu^- X$	$W^* \bar{s} \to \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
$e^{\pm} p \to e^{\pm} X$	$\gamma^* q \to q$	$g, q, \bar{q}$	$0.0001 \lesssim x \lesssim 0.7$
$e^+ p \to \bar{\nu}\, X$	$W^+\{d,s\} \to \{u,c\}$	d, s	$x \gtrsim 0.01$
$e^{\pm} p \to e^{\pm} c\bar{c}\, X$	$\gamma^* c \to c$, $\gamma^* g \to c\bar{c}$	c, g	$0.0001 \lesssim x \lesssim 0.01$
$e^{\pm} p \to \text{jet} + X$	$\gamma^* g \to q\bar{q}$	g	$0.01 \lesssim x \lesssim 0.1$
$p\bar{p} \to \text{jet} + X$	$gg, qg, qq \to 2j$	g, q	$0.01 \lesssim x \lesssim 0.5$
$p\bar{p} \to (W^{\pm} \to \ell^{\pm}\nu)\, X$	$ud \to W, \bar{u}\bar{d} \to W$	$u, d, \bar{u}, \bar{d}$	$x \gtrsim 0.05$
$p\bar{p} \to (Z \to \ell^+\ell^-)\, X$	$uu, dd \to Z$	d	$x \gtrsim 0.05$

Table 1.4.: List of processes included in MSTW fits. Taken from [4].

sections, the coefficient functions are directly included at the appropriate order in the fitting code, however for the case of jet cross sections this is done through the FastNLO package which is described in detail later in this chapter.

The set of parameters which provides the best global fit to all data is the central fit. The χ^2 values for each data set from the central fit are given in Table 1.5.

Once the central value is found, the PDF errors are evaluated using the Hessian method. This is based on the assumption that the global fit is quadratic in variations of each PDF parameter about the minimum, i.e.

$$\chi^2_{global} - \chi^2_{min} = \sum_{i,j=1}^{n} H_{ij}(a_i - a_i^0)(a_j - a_j^0) \tag{1.25}$$

where the PDF parameters are $\{a_1, ..., a_n\}$ with best fit points $\{a_1^0, ..., a_n^0\}$, and the matrix H is:

Data Set	LO	NLO	NNLO
BCDMS μp F_2 [10]	165/153	182/163	170/163
BCDMS μd F_2 [11]	162/142	190/151	188/151
NMC μp F_2 [13]	137/115	121/123	115/123
NMC μd F_2 [13]	120/115	102/123	93/123
NMC $\mu p/\mu d$ [14]	131/137	130/148	135/148
E665 μp F_2 [15]	59/53	57/53	63/53
E665 μd F_2 [15]	49/53	53/53	63/53
SLAC μp F_2 [16], [17]	24/18	30/37	31/37
SLAC μd F_2 [16], [17]	12/18	30/38	26/38
NMC/BCDMS/SLAC F_L [10]- [12]	28/24	38/31	32/31
E866/NuSea pp DY [18]	239/184	228/184	237/184
E866/NuSea pd/pp DY [19]	14/15	14/15	14/15
NuTeV νN F_2 [20]	49/49	49/53	46/53
CHORUS νN F_2 [21]	21/37	26/42	29/42
NuTev νN xF_3 [20]	62/45	40/45	34/45
CHORUS νN xF_3 [21]	44/33	31/33	26/33
CCFFR $\nu N \to \mu\mu X$ [22]	63/86	66/86	69/86
NuTeV $\nu N \to \mu\mu X$ [22]	44/40	39/40	45/40
H1 MB 99 e^+p NC [23]	9/8	9/8	7/8
H1 MB 97 e^+p NC [24]	46/64	42/64	51/64
H1 low Q^2 96-97 e^+p NC [24]	54/80	44/80	45/80
H1 high Q^2 98-99 e^-p NC [25]	134/126	122/126	124/126
H1 high Q^2 99-00 e^+p NC [26]	153/147	131/147	133/147
ZEUS SVX 95 e^+p NC [27]	35/30	35/30	35/30
ZEUS 96-97 e^+p NC [28]	118/144	86/144	86/144
ZEUS 98-99 e^-p NC [29]	61/92	54/92	54/92
ZEUS 99-00 e^+p NC [30]	75/90	63/90	66/90
H1 99-00 e^+p CC [31]	28/28	29/28	29/28
ZEUS 99-00 e^+p CC [32]	36/30	38/30	37/30
H1/ZEUS ep F_2^{charm} [33]- [39]	110/83	107/83	95/83
H1 99-00 e^+p incl. jets [40]	109/24	19/24	-
ZEUS 96-97 e^+p incl. jets [41]	88/30	30/30	-
ZEUS 98-00 $e^\pm p$ incl. jets [42]	102/30	17/30	-
D0 II $p\bar{p}$ incl. jets [43]	193/110	114/110	123/110
CDF II $p\bar{p}$ incl. jets [44]	143/76	56/76	54/76
CDF II $W \to l\nu$ asym. [45]	50/22	29/22	30/22
D0 II $W \to l\nu$ asym. [46]	23/10	25/10	25/10
D0 II Z rap. [47]	25/28	19/28	17/28
CDF II Z rap. [48]	52/29	49/29	50/29
Total	3066/2598	2543/2699	2480/2615

Table 1.5.: Table of χ^2 values for each data set included in the MSTW2008 fit at LO, NLO and NNLO. Taken from [4].

$$H_{ij} = \frac{1}{2}\frac{\partial^2 \chi^2_{global}}{\partial a_i \partial a_j} \tag{1.26}$$

From this, the uncertainty originating from PDFs on any physical quantity can be written as:

$$\Delta F = T\sqrt{\sum_{i,j=1}^{n} \frac{\partial F}{\partial a_i} C_{ij} \frac{\partial F}{\partial a_j}} \tag{1.27}$$

where C is the covariance matrix defined as the inverse of the Hessian matrix, and T is the tolerance $T = (\Delta\chi^2_{global})^{1/2}$, which is defined as the increase in χ^2 away from the global minimum for the required confidence level.

It is easier in practice to find the eigenvectors of the covariance matrix and use variations in this space to define PDF errors, as first shown in [9]. The covariance matrix, due to being symmetric, has an orthonormal set of eigenvectors, with each corresponding to a parameter in the PDF fit. From these eigenvectors e_{ik}, displacements from the best fit parameter set can can be written as:

$$a_i - a_i^0 = \sum_{k=1}^{n} e_{ik} z_k. \tag{1.28}$$

From this, individual PDFs corresponding to movements in each eigenvector direction can be defined. All parameters are left at the global minimum except the one in question, which is moved according to a 68% or 90% tolerance level. This is represented by the following equation:

$$a_i(S_K^{\pm}) = a_i^0 \pm t e_{ik}. \tag{1.29}$$

where t is a coefficient to ensure the the required tolerance condition $T = (\Delta\chi^2_{global})^{1/2}$ is met. Each eigenvector now corresponds to a pair of eigenvector PDFs, one moving in each direction from the global minimum. Now, instead of calculating the correlation matrix for each observable for which a PDF error must be calculated, one can simply calculate the change in the observable for each eigenvector PDF set $S^{\pm}_K$, and sum over these deviations in quadrature:

$$\Delta F = \frac{1}{2}\sqrt{\sum_{i,j=1}^{n}[F(S_k^+) - F(S_k^-)]^2}. \tag{1.30}$$

where the uncertainties are assumed to be symmetric about the central value. It is possible to include antisymmetric uncertainties for such cases. The tolerance value, T, is itself determined within the MSTW method by using a dynamic method. This is necessary due to the fact that the included data sets may not be consistent, and the errors are not always Gaussian about the global fit. The method used is to determine the 68% and 90% confidence levels for each individual data set under variations in each eigenvector. The 68% and 90% tolerance level for any given eigenvector is then defined as the distance one can move from the global fit in its direction whilst keeping all data sets within their respective confidence levels.

The 2008 MSTW PDFs are shown at two different scales in Fig 1.4 along with their uncertainties. Clearly the gluon comes to dominate as the scale increases, due to the prevalence of quark-gluon and gluon-gluon splittings at low x. At LHC energies, gluons constitute the majority of interactions for many processes. There are some observables, such as W/Z production, which may only be mediated by quarks, however the evolution of these distributions are governed very strongly by the gluon distribution.

1.4. The Theory of Hadronic Jets

When considering observable quantities which involve the particles of the strong force, one must again consider the effect of colour confinement. Whilst hard interactions between hadrons are mediated by free quarks and gluons, these particles are not

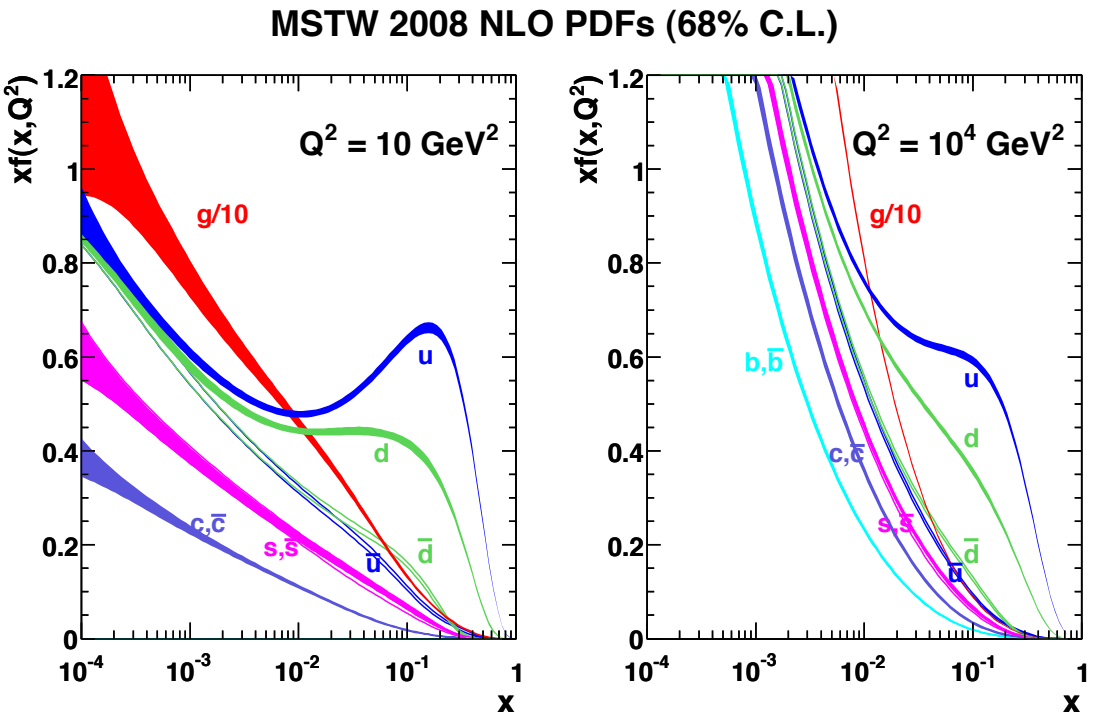


Figure 1.4.: The MSTW 2008 NLO PDFs at two different evolution scales, showing the dominance of the gluon as center of mass energy increases. Plot taken from [4].

physically observable. At larger distance scales, only colourless bound states may be detected. In order to bridge this gap between the hard partonic interactions and the soft final state, one must model the evolution of parton branchings.

For instance, a process described by the QCD matrix element with two final state gluons will undergo soft radiation in both the initial and final state, such that many colourless hadrons will be observed in the detector. These bunches of hadrons are named jets, and the dynamics of these objects must be understood in order to obtain information from the hard scatter.

The most simple definition of a jet is that of the cone algorithm [49], in which a jet is defined as the particles that flow into a cone of set distance in η, ϕ space and above a certain energy threshold. Cone algorithms must be constructed carefully to ensure infrared safety. This means that the emission of a soft or collinear parton will not alter the physical prediction, an effect which would otherwise render calculations incredibly unstable in the non-perturbative region.

Another definition, which is used at HERA and most recently the LHC, is based on sequentially recombining final state hadrons until a threshold is reached, at which point the collected particles are defined as a single jet. These are called clustering algorithms [50], and several variations can be constructed. The threshold definition in the general case is:

$$d_{ij} = min(k_{T,i}^{2p}, k_{T,j}^{2p}) \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R} \tag{1.31}$$

There are three instances of k_T power choice: $p = -1, 0, 1$. 1 corresponds to the k_T clustering algorithm, 0 corresponds to Cambridge-Aachen algorithm, and -1 is the anti-k_T algorithm [51]. For jet cross sections at the Tevatron, the k_T algorithm was sometimes chosen as well as the standard cone algorithm, whereas the LHC uses anti-k_T. This algorithm is the newest, only having come into recognition in around 2007 for the LHC. The benefit of using this over the k_T algorithm is the resulting jets appear very cone-like, so the benefits of cone algorithms are maintained with the guarantee of infrared safety from a recomination algorithm. The parameter R is a variable which can be varied and should not affect the physics in a meaningful manner. For ATLAS jet analyses, two choices of $R = 0.4$ and $R = 0.6$ were simultaneously used [52], whilst CMS uses $R = 0.7$ [53].

1.5. Theoretical Tools

In order to produce general theoretical cross sections, one must accurately model both the large momentum transfer interactions which occur at the collision point and also the soft radiation which occurs both before and after the interaction. There are many computer programs which have been publicly produced for the particle physics community. These are all based around a common concept of calculating matrix elements using Quantum Field Theory for the hard interactions, and dressing these matrix elements with soft radiation known as parton showering. This section will summarize the principles of, and differences between, the tools which have been used in this thesis.

1.5.1. Herwig++

Herwig++ [54] is a Monte Carlo generator for many processes in particle physics. Specifically, for DIS it incorporates neutral- and charged-current electroweak processes between leptons and hadrons.

Initial- and final-state radiation is simulated through parton showering, which can approximate the numerous QCD emissions involved in hadron interactions. The starting point is the assumption that the emission process sharply diverges for soft and collinear radiation. Consider the emission of a single gluon from a final-state quark, the amplitude for which is governed by the propagator:

$$\frac{1}{(p_q + p_g)^2} \approx \frac{1}{2E_q E_g(1 - \cos\theta_{qg})}$$

Clearly the two cases of $\cos\theta_{qg} \to 1$ and $E_{q/g} \to 0$ dominate, and so parton emissions will tend to be clustered in the direction of the final state hard parton. In order to describe each emission in a shower separately, one can consider the subsequent branchings as an individual current which factorises from the splitting in question. By also assuming the emitted gluon to be almost on shell, and summing over helicities and polarisations, one can derive the amplitude for a single splitting $P_{ij}(z)$. These are the QCD splitting functions representing the probability of a single off-shell parton splitting into two daughter partons carrying momentum fractions z and $1 - z$ of the parent's momentum. There are 4 QCD splitting functions which describe the possible configurations: $q \to qg, q \to gq, g \to qq, g \to gg$, shown in Table 1.3.

From these assumptions, one can model the initial and final state radiation using a Sudakov Form Factor, $\Delta_i(\mu^2)$, which represents the probability of no emission between two scales μ_0 and μ:

$$\Delta_i(\mu^2) \equiv \exp\left[-\sum_j \int_{\mu_0^2}^{\mu^2} \int dz \frac{\alpha_s}{2\pi} P_{ij}(z)\right]$$

For final-state radiation, the shower is started at the hard scale of the scattering process and continues until a pre-determined scale μ_f, is reached at which point hadronisation is applied. For initial-state particles, a backward evolution scale is

used, which again starts at the hard-scattering level until the parton can be treated as part of a PDF.

By using this approximation, Herwig++ can model the very complex soft processes which are inherent in hadron colliders without the necessity of calculating the matrix elements.

Once the initial-state showering has produced a hard parton to be included in the scattering, Herwig++ calculates the matrix element for the required process. For the case of neutral current DIS, only the LO diagram to zero order in α_s, but first order in α_{EM}, is included. This represents a problem for modelling jet production, since the processes of boson-gluon fusion and QCD Compton, which are both order α_s^2 are minimally required to produce hard back-to-back jets.

These processes are first order in jet production, but are $O(\alpha_s)$ in QCD, and are not included in the Herwig++ library of matrix elements. The only possibility for Herwig++ to generate jets is via the parton shower scheme, which allows reasonably high emissions with a small probability. This leads to the expectation that lower energy jets which do in fact originate from QCD radiation will be modelled by Herwig++, but any hard jets which are produced at the parton level will not be produced.

1.5.2. Sherpa

Sherpa [55] is an independent Monte Carlo program which implements its own method of parton showering and hadronisation. The problem of higher-order corrections to the matrix element is tackled through the Matrix Element-Parton Shower merging scheme.

The idea of this scheme is to choose a cut in the emission phase-space, Q_{cut}, which defines the point at which the hard partonic subprocess ends and the parton shower scheme begins. This is usually a k_T jet measure. Above Q_{cut}, additional hard emissions are allowed in the matrix element up to a maximum predefined multiplicity. This allows well separated hard partons to be created, whilst the soft and collinear emissions are in the domain below Q_{cut} in the form of a usual parton shower.

Sherpa makes use of the CKKW algorithm [56] to perform the merging between the matrix element (ME) region and the parton shower (PS) region. The hard partons are produced by the in-built generator, AMEGIC++, up to a given multiplicity N_{jet}. Naively, one would suggest to simply take these final state partons and dress them with PS radiation to produce an N_{jet}-jet final state. This cannot be done, however, due to the fact that the merging parameter Q_{cut} is not a physical quantity and so no residual dependence upon it can be kept. This is discussed within the context of DIS in Chapter 2.

1.5.3. POWHEG

POWHEG [57] is a method of interfacing NLO matrix elements with parton shower algorithms. Unlike Sherpa, POWHEG includes the full NLO matrix element in its calculations, and so any parton shower effects will simulate effects at higher orders than NLO. For the studies in this thesis, the POWHEG-BOX program was used, which allows the NLO parton level events to be interfaced with any parton shower Monte Carlo program. In this case, the Herwig++ shower can be used as the interface in order to directly compare the effects of showering with different order matrix elements.

1.5.4. NLOJet++

NLOJet++ [58] [59] is a tool to calculate jet cross sections to full NLO accuracy in QCD for ep, $p\bar{p}$ and pp collisions. The code utilises the Catani-Seymour subtraction method to regulate and remove the ultraviolet divergences which appear in the cross sections.

To calculate a cross section with m final state partons, the calculation takes the QCD matrix elements, and calculates the integral:

$$\sigma = \int_m d\sigma^B + \int_{m+1} d\sigma^R + \int_m d\sigma^V \tag{1.32}$$

where σ^B is the LO Born cross section, σ^R is the real emission correction and σ^V is the virtual correction. The final two terms constitute the NLO correction, which both contain ultraviolet divergences which cancel those encountered in the Born term.

Whilst this integral can be and has been performed analytically for many basic processes, the key to the Catani-Saymour method is to create a general form which can be integrated using a partonic Monte Carlo program. This is achieved by rewriting the NLO part of the calculation with a counterterm σ_A

$$\sigma_{NLO} = \int_{m+1} [d\sigma^R - d\sigma^A] + \int_{m+1} d\sigma^A + \int_m d\sigma^V \tag{1.33}$$

If the counterterm has the same singular behaviour as $d\sigma_R$, then the left hand term can be integrated numerically, and the singular behaviour is contained in the final two terms. The only analytical calculation that needs to be performed is that of $d\sigma_A$ over the additional one-parton phasespace, then the poles can be combined with those of $d\sigma_V$, and the only remaining calculation required is to generated weighted partonic events using a Monte Carlo program for events with m final state partons. This is the role of NLOjet++, which has been shown to agree excellently with hadron collider QCD results.

1.5.5. FastNLO & APPLgrid

Whilst NLOjet++ is an excellent tool for calculated full NLO pQCD cross sections, it requires computing time on the order of weeks to obtain the statistical precision necessary for comparisons to data. This is fine for individual calculations, but if one wishes to perform a full PDF analysis, the calculation must be made for many different PDFs and scale choices. For example, in the MSTW PDF fitting procedure, the χ^2 must be calculated every time the PDF parameters are changed, in order to determine the gradient in the parameter space.

This leads to the problem that the calculation times required for a PDF fit are unworkable, and so a technique for quickly changing the input PDF for a given

NLOjet++ calculation is required. Two methods have been created for this which rely on similar principles: FastNLO [60] and APPLgrid [61].

FastNLO separates the α_s and PDF contributions from the partonic subprocess calculations, and introduces an interpolation function to obtain the x and $\mu_{F,R}$ dependence of these pieces. The result is to reduce the total cross section:

$$\sigma(\mu_r, \mu_f) = \sum_{n,i} c_{n,i}(x, \mu_r, \mu_f) \otimes [\alpha_s^n(\mu_r) \cdot f_i(x, \mu_f)] \tag{1.34}$$

to an approximate product given by

$$\sigma(\mu) \approx \sum_{n,i,k,l} \tilde{\sigma}_{n,i,k,l}(\mu) \cdot \alpha_s^n(\mu^m) \cdot f_i(x^k, \mu^l). \tag{1.35}$$

where the NLOjet++ is only performed once to produce the partonic cross section $\tilde{\sigma}_{n,i,k,l}$ at arbitrary points in PDF and α_s phase space.

APPLgrid uses an effectively identical process in which the NLOjet++ run is used to produce a grid of weights in (x, Q^2) space, which can then be similarly convoluted with an arbitrary PDF. A higher order interpolation between grid points is used for APPLgrid than FastNLO which allows a better tradeoff between accuracy and memory use. A key difference between the two programs in the context of this thesis is the inclusion of user-customisable modules, so new grids may be produced by the user. This has allowed for a more detailed study of different scale choices in relation to MSTW PDFs to be made with APPLgrid, as opposed to FastNLO where the scale choice is fixed by the pre-produced tables.

Chapter 2.

Description of DIS Jet Cross Sections with Modern Monte Carlo Generators

A process of importance to PDF fitting is that of inclusive jets in DIS. With only one initial state hadron, a full understanding of final state jets can provide a great deal of information on the underlying QCD. These cross sections, from both ZEUS and H1, are included for neutral current processes in MSTW fits, using NLO pQCD calculations provided by FastNLO. No charged current jet production cross sections are included in any PDFs to date, although it is possible that these cross sections could provide unique information on the down quark distribution.

This chapter will outline two studies conducted into such cross sections. First, the non-perturbative effects on the neutral current inclusive jet cross section are investigated by producing calculations with the newest available Monte Carlo generators. Secondly, the charged current inclusive jet cross section is investigated as a potential window into the down quark PDF.

2.1. Neutral Current Inclusive Jets

The neutral current inclusive jet cross section from ZEUS [62], provides information on both the gluon distribution and α_s, and is a key component of the MSTW PDF analysis. Predictions for the PDF fits are produced using NLOjet++ within FastNLO,

and as such do not directly include hadronisation or parton showering effects. The calculated NLO correction is much larger than the inclusive DIS total cross section or proton-(anti)proton induced inclusive jets, and as such the scale variations are relatively large and shape-dependent. This study will use various LO Monte Carlo generators to study the nature of the higher order effects, especially at low E_T, where the soft parton shower effects are largest.

The analysis used for the jet cross section in question did not already exist in the current Rivet [63] analysis software release, and so it was written as a user-created plug-in analysis. The code required to obtain the final state particles in the Breit frame was also missing from the software, and so was included by hand. The Breit frame is defined as the frame in which the the energy of the lepton is the same before and after the collision. It is used in this cross section due to the property of isolating the hadronic final state particles from the rest of the event.

The selection criteria were taken from the paper [62] and are as follows:

- $Q^2 > 125\text{GeV}^2$
- Jets of $E_T > 8\text{GeV}$ and $-2 < \eta < 1.8$
- $-0.7 < \cos(\gamma) < 0.5$

where γ is the angle of the scattered quark, calculated through the Jaquet-Blondel definition of the DIS variables:

$$y_{JB} = \frac{\sum_h (E - p_z)_h}{2E_e}, \quad Q^2_{JB} = \frac{p^2_{T,h}}{1 - y_{JB}}. \tag{2.1}$$

2.1.1. Event Generator Predictions

The events generated were specified according to the same cuts detailed above. The results for the differential jet cross-sections with respect to Q^2 and E_T are given in Fig 2.1.

Both the $\frac{d\sigma}{dQ^2}$ and $\frac{d\sigma}{dE_T}$ plots show that the Monte Carlo is on the whole below the data, which is to be expected from purely LO matrix elements. The matrix elements used in Herwig++ contain only the diagrams involving a quark directly

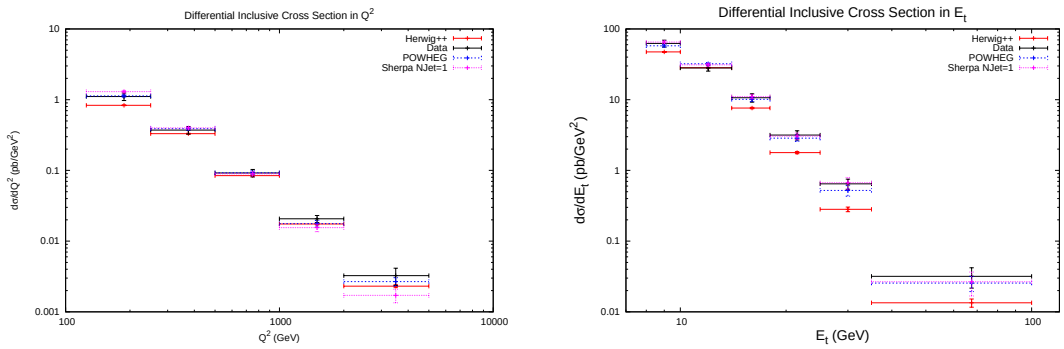


Figure 2.1.: Differential jet cross-sections generated by Herwig++ compared to ZEUS data.

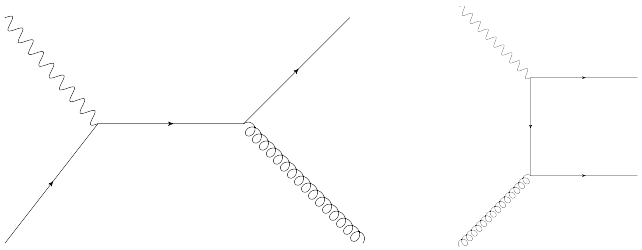


Figure 2.2.: Diagrams for QCD Compton and Boson-Gluon Fusion processes

interacting with the exchanged photon. It therefore does not directly include the QCD Compton and Boson-Gluon Fusion processes, shown in Fig 2.2, which will contribute to jet production, especially at low Q^2. To compensate for the lack of these matrix elements, additional contributions from the parton shower are required. The breakdown of the $\frac{d\sigma}{dE_T}$ cross section into Q^2 bins is shown in Fig 2.3 alongside the Sherpa and Herwig++ predictions. The results indicate that at low Q^2, the parton shower does not replicate the required matrix elements to produce enough high E_T jets.

The E_T dependence is further studied by splitting the events into 6 different plots of increasing Q^2 transfer. These plots are shown in Fig 2.4. For the lower Q^2 bins, it is clear that Herwig++ produces far fewer hard jets than is to be expected, whereas the higher Q^2 plots show an improvement in this respect, albeit with a drop in the lower E_T jets. POWHEG, on the other hand, does not suffer these issues at low Q^2, and fills out the E_T spectrum in all of the plots. Clearly the inclusion of the NLO

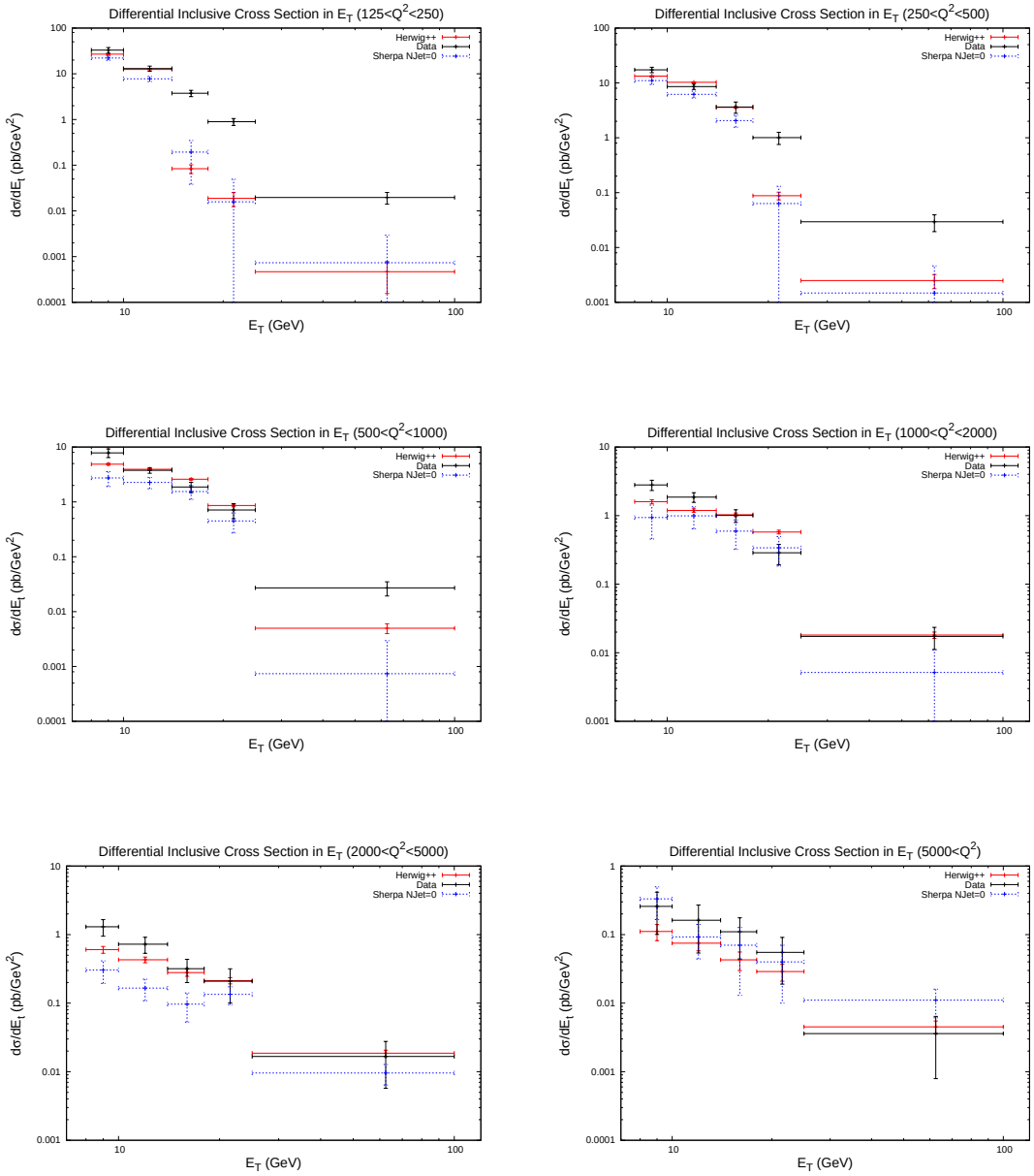


Figure 2.3.: Comparison between Herwig++ and Sherpa with no extra emissions at the parton level.

matrix element with the same showering as Herwig++ is enough to prevent this deficit from occuring.

The same kinematical values were used to generate events through Sherpa as with Herwig++ previously, and the events were run through the same Rivet analysis.

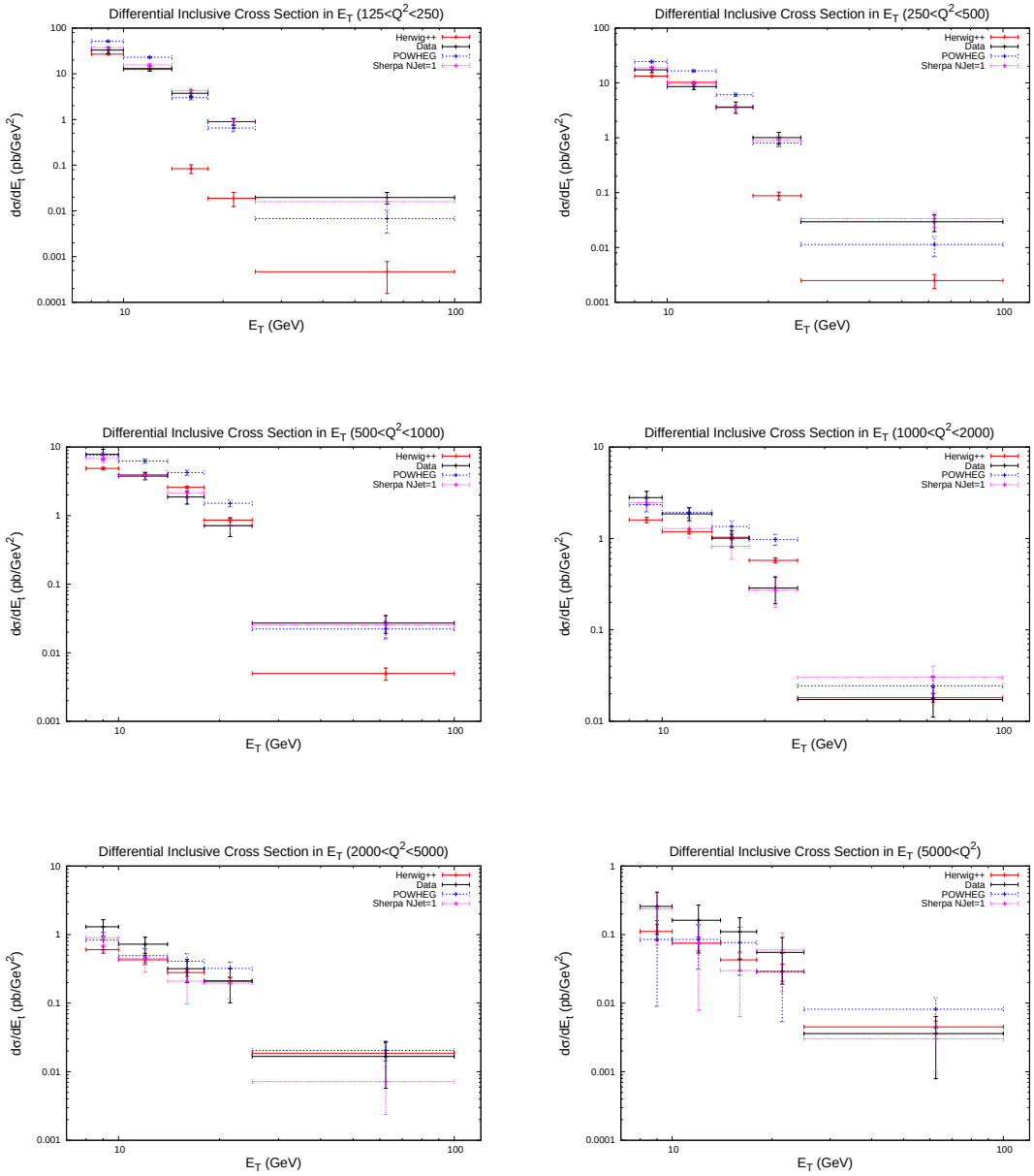


Figure 2.4.: Comparison between Herwig++, Powheg and Sherpa with one extra emission at the parton level.

One variable which is present in Sherpa that is not in Herwig++, however, is the parameter N_{jet}, which is the maximum number of additional emissions allowed from the hard process.

For example, the $N_{jet} = 1$ modification to a $q\bar{q}$ final-state cross-section is calculated from the diagram in Fig 2.5:

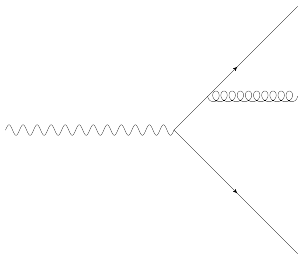


Figure 2.5.: ($N_{jet} = 1$) modification to a 2-parton final-state

$$\sigma_3^{(ME)}(Q_{cut}) = \sigma_2 \cdot 2[\Delta_q(Q, Q_{cut})]^2 \int_{Q_{cut}}^{Q} dq \Gamma_{q\to qg}(q) \Delta_g(q, Q_{cut}) \tag{2.2}$$

where σ_i is the i parton final-state cross-section and $\Gamma_{q\to qg}(q)$ is the probability for the given branching to take place. The $\Delta_{q/g}$ are the quark and gluon Sudakov form factors, which represent the probability of the relevant parton to evolve between the two scales in the argument without emitting any resolvable radiation. This can be seen to describe two quarks evolving from Q_{cut} to Q with the emission of a gluon at some scale q which is integrated over. The integral is weighted by the matrix element for the emission process.

Similar expressions can be obtained for any number of final state partons, by simply considering the possible combinations of splittings within the diagram and weighting the ME with suitable Sudakov factors.

This must now be merged with a parton shower in such a way as to cancel the dependence on Q_{cut} in order to obtain a sensible jet cross-section. This is done by considering that the total 3-jet cross-section will be a combination of a three parton final state followed by a parton shower with no new resolvable jets produced, and a two parton matrix element followed by a resolvable jet emission in the parton shower. This is written as:

$$\begin{aligned}\sigma_3^{TOT}(Q_0) &= \sigma_3^{(ME)}(Q_{cut})[\Delta_q(Q_{cut},Q_0)]^2\Delta(Q_{cut},Q_0) \\ &\quad +\sigma_2\times 2[\Delta_q(Q_{cut},Q_0)]^2\int_{Q_0}^{Q_{cut}} dq\alpha_s(q)\Gamma_{q\to qg}(q)\Delta_g(q,Q_0),\end{aligned}$$

and substituting in for Eq 2.2 gives the total 3-jet cross-section as

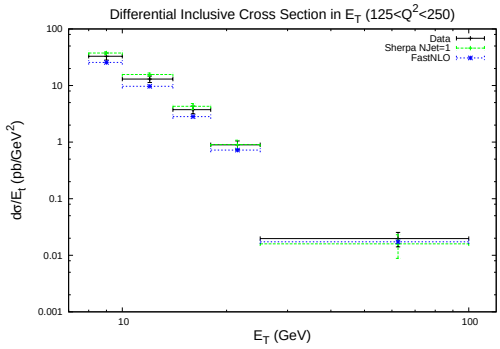
$$\sigma_3^{TOT}(Q_0) = \sigma_2\times 2[\Delta_q(Q_{cut},Q_0)]^2\int_{Q_0}^{Q} dq\alpha_s(q)\Gamma_{q\to qg}(q)\Delta_g(q,Q_0). \tag{2.3}$$

This is independent of Q_{cut} to NLL, as required. By using this technique, Sherpa can generate jet cross-sections to any multiplicity, and so can mimic higher-order processes without making any NLO computations.

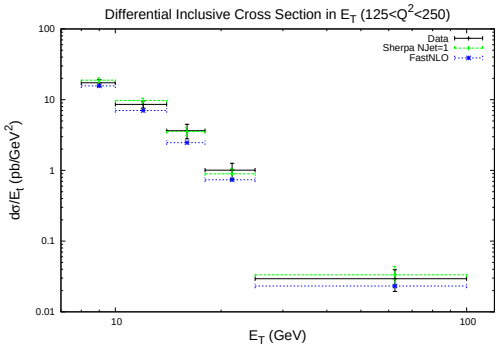
Since Herwig++ contains only the leading order diagram, it is expected that by setting $N_{jet} = 0$, a similar result should occur. Similarly, $N_{jet} = 1$, which corresponds to LO diagrams with one extra emission, should correspond to FastNLO at LO (in α_s). The full NLO implementation of FastNLO is not used in this comparison because this would be to $\mathcal{O}(\alpha_s^2)$, and contain two-emission tree diagrams, as well as loop corrections to the one-emission cross-sections which are not included in Sherpa.

The comparison to Herwig++ is shown in Fig 2.3 and to FastNLO in Fig 2.6. It is clear that the $N_{jet} = 0$ case contains the same issues as Herwig++, with a deficit of high E_T jets at low Q^2, although the two do not agree numerically. The agreement is somewhat better for FastNLO with Sherpa, despite the fact that the FastNLO data is purely parton-level.

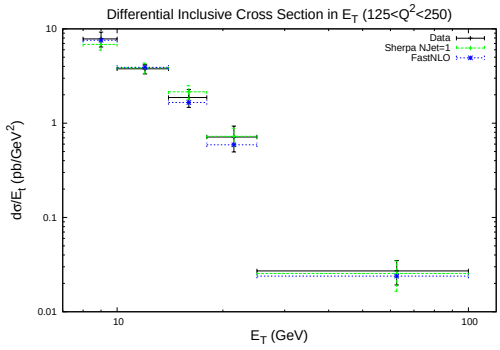
In order to further see where the differences between Herwig++ and Sherpa are arising, profile plots of the variable Q^2, x, y and E_e (energy of the recoiling lepton) are produced. These plots are shown in Fig 2.7. The profiles demonstrate that the kinematics of Sherpa and POWHEG generally agree, while Herwig++ deviates from the others. There is a deficit of events at low Q^2 and high inelasticity for Herwig++, whilst the scattered lepton peak is narrower and taller. This explains the inability for Herwig++ to produce high-p_T jets, especially at low Q^2. The matrix element with showering does not produce hard enough emissions to describe the data. Sherpa and



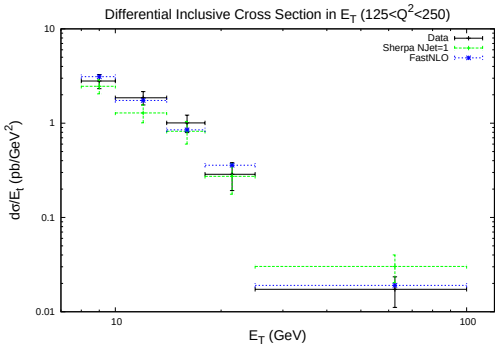
(a) $125 < Q^2 < 250$



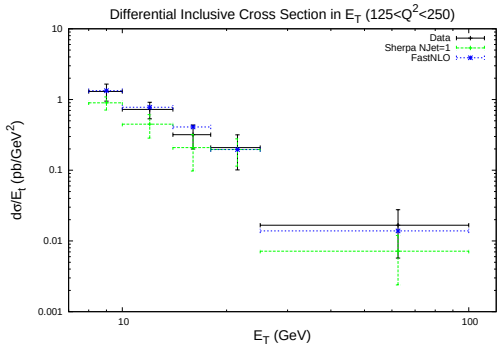
(b) $250 < Q^2 < 500$



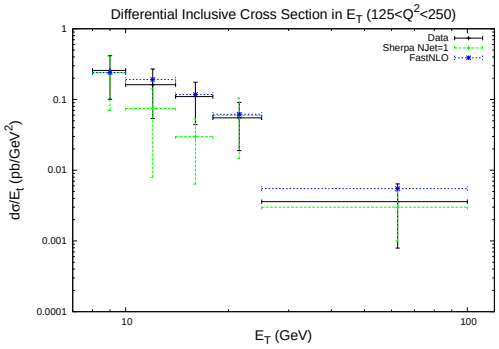
(c) $500 < Q^2 < 1000$



(d) $1000 < Q^2 < 2000$



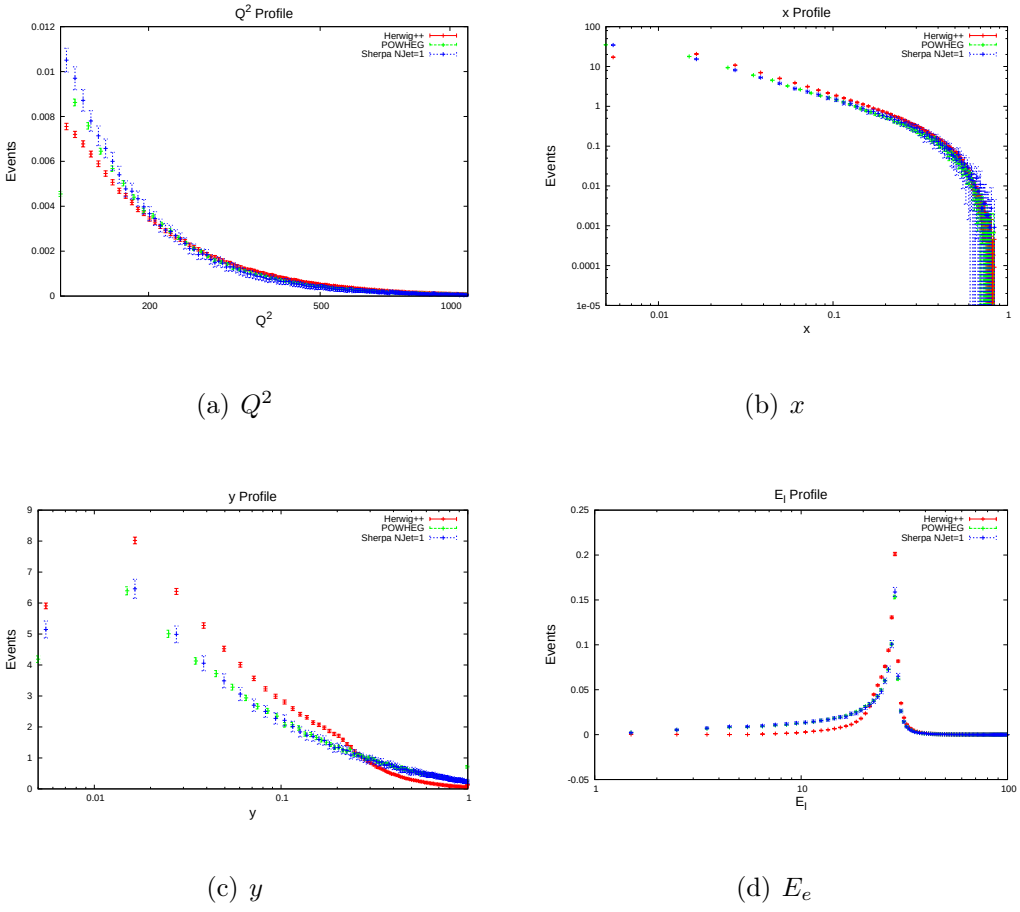
(e) $2000 < Q^2 < 5000$



(f) $5000 < Q^2$

Figure 2.6.: Comparison between FastNLO and Sherpa with one extra emission at the parton level.

POWHEG, however, with their matrix element-parton shower matching methods, can fill the softer region of phasespace at low Q^2. The fact that Sherpa is generating events with higher inelasticity and lower positron energy shows that it is better suited to producing high energy jets.



(a) Q^2 (b) x

(c) y (d) E_e

Figure 2.7.: Comparison between Herwig++ and Sherpa for the profiles of the main DIS variables.

The kinematic plots demonstrate the role of parton showering in replicating jet data. Whilst the lack of inclusion of the BGF and QCDC processes in Herwig++ led to a deficit of events in the NC cross section, this is not an important effect here. At high Q^2, Herwig++'s lack of hard final state partons in the matrix element is compensated by the fact that the parton shower is more likely to produce a hard jet with a harder starting scale because of the higher starting scale in the evolution.

2.2. ZEUS charged current Jets

An observable which has potential impact on an otherwise relatively unconstrained PDF is charged current jet production. Such cross sections have never been included in PDF fits; although charged current DIS is considered as a total cross section in the MSTW fits, the jet cross sections are not. Here the ZEUS charged current jet cross section in positron-proton DIS [64] is studied.

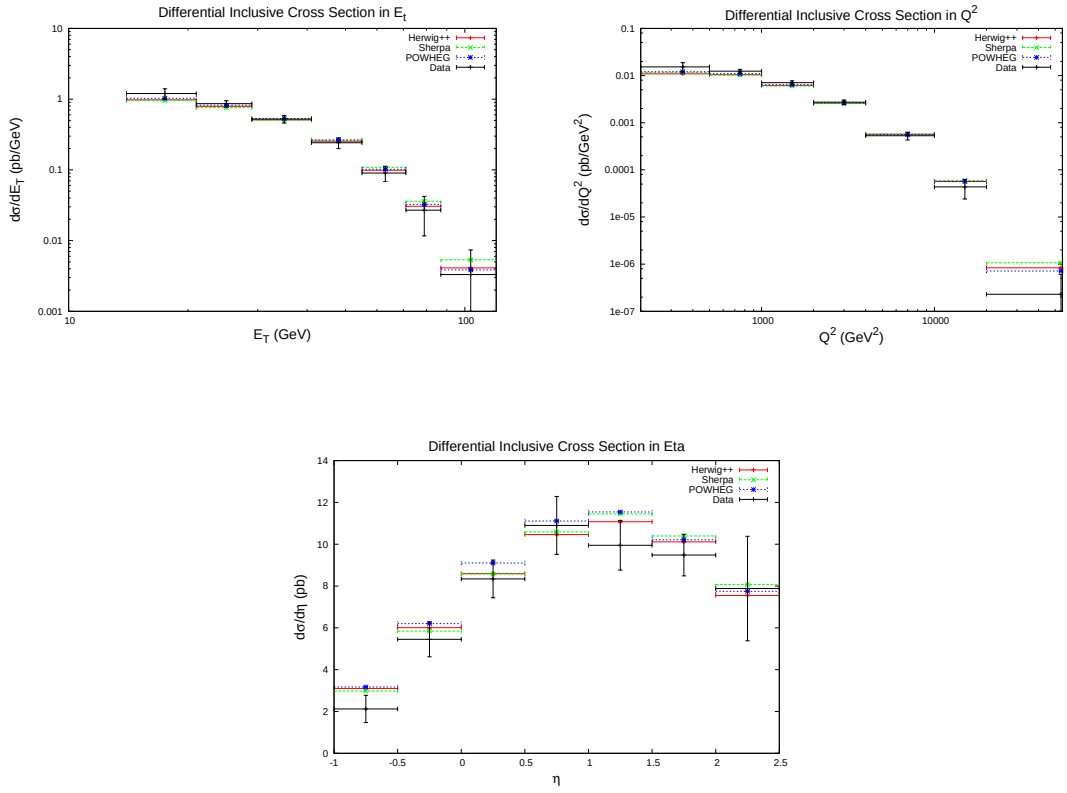


Figure 2.8.: Charged current inclusive jet cross sections at ZEUS as a function of E_T, Q^2 and η respectively.

The differential cross sections binned in Q^2, E_T and η are shown in Fig 2.8. Again, the predictions are calculated using Herwig++, Sherpa and POWHEG. For this observable, all of the predictions are in reasonable agreement with the data, with the exception of Sherpa, which gives a prediction higher than data for some points in the phasespace. Herwig++ does not have the same problems as in the neutral current cross section. This is due to the higher Q^2 cut in the charged current process.

All events with $Q^2 < 200GeV$ are cut from this analysis, although relatively few should occur below this limit due to supression of Q^2 below the W mass. This means we are in the region which Fig 2.7(a) shows the generators agree on the underlying kinematics, and the lack of a hard vertex in the matrix element is not as important as for the neutral current jets.

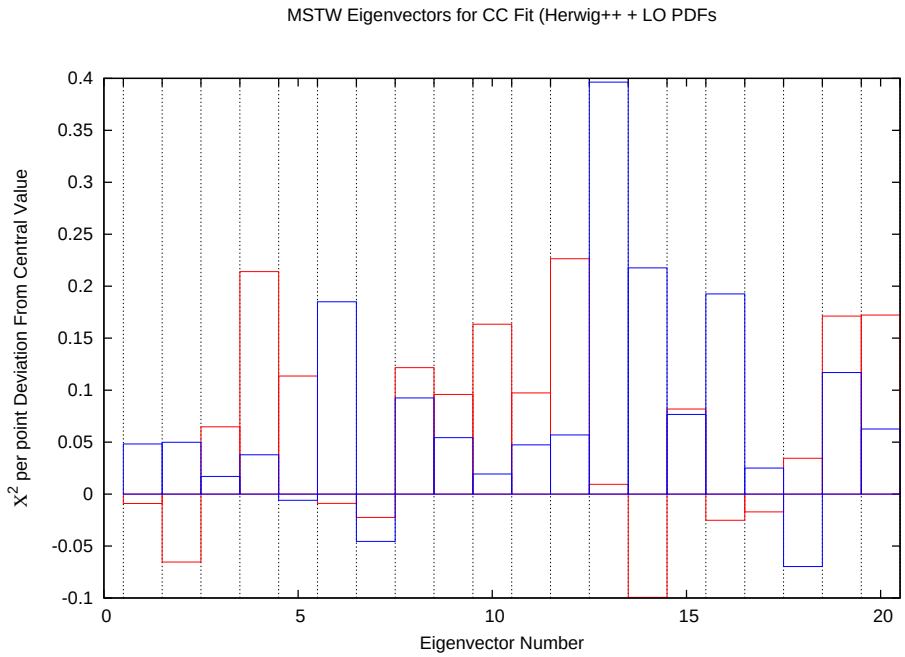


Figure 2.9.: Change in fit quality for each eigenvector in the MSTW 2008 LO set for Herwig++ CC inclusive jet predictions.

In order to test the potential constraining of the down quark using the dataset, the calculation is made for each of the standard MSTW eigenvectors. Due to the excellent agreement from Herwig++ and POWHEG, these generators are used for the calculations. LO PDFs are used for the Herwig++ calculations, and NLO PDFs are used for POWHEG.

To quantify the changes in the cross section with movements in the space of MSTW eigenvector sets, a goodness of fit measure must be defined. For this measurement, the total systematic errors summed in quadrature are in general smaller than the statistical errors, and so a simple definition is used, summing the statistical and total systematic errors in quadrature. The χ^2 measure is defined by:

$$\chi^2 = \frac{(D_i - T_i)^2}{\sigma_q^2} \tag{2.4}$$

where D_i, T_i are the data and theory points respectively, and σ_q is the sum in quadrature of the statistical and systematic uncertainties. The sensitivity to the data to a particular input PDF can be observed by testing for significant changes in the fit quality along a particular eigenvector direction. The results for the Herwig++ and POWHEG calculations are shown in Figs 2.9 & 2.10. These plots demonstrate how each eigenvector effect the fit to data, by showing the change in χ^2 which occurs when moving 1 σ in either direction along the relevant eigenvector. Therefore, the eigenvectors which are best suited to the data are those with the largest bars in the negative direction..

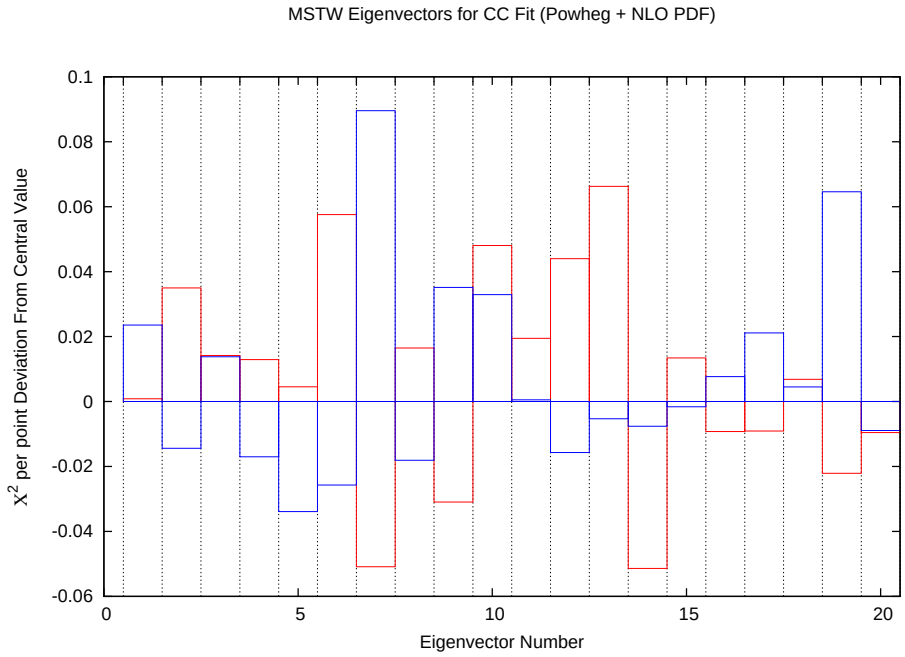
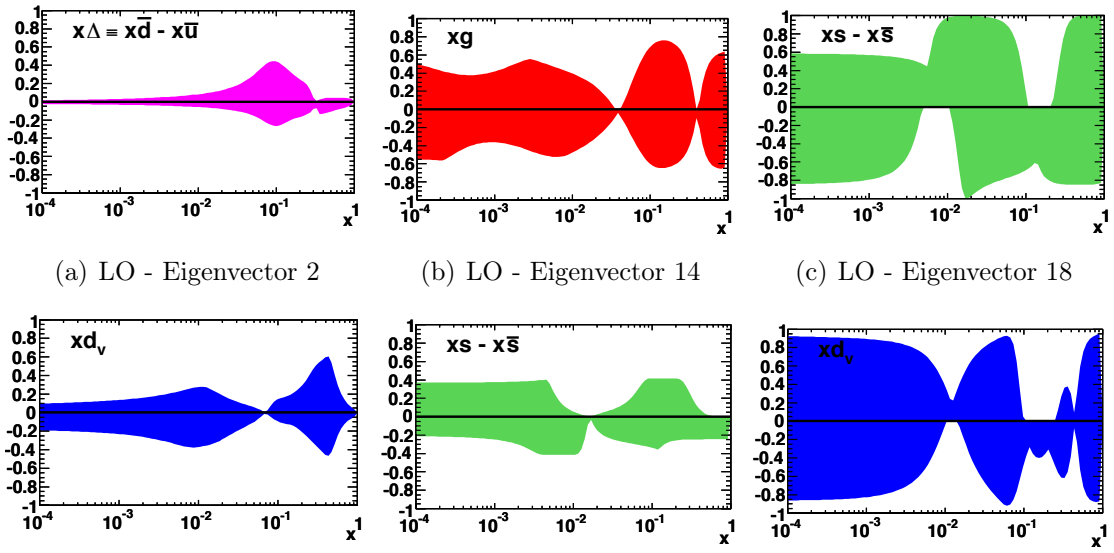


Figure 2.10.: Change in fit quality for each eigenvector in the MSTW 2008 NLO set for POWHEG CC inclusive jet predictions.

The LO predictions display little favourable dependence on the PDFs, with very few eigenvectors contributing any improvement in fit at all. Most of the eigenvectors

show a reduction in fit quality, with only eigenvectors 2, 14 and 18 showing a noticeable improvement in the χ^2, which improvements in fit quality of $0.05 - 0.1$ for each. The first of these is affected mostly by the $(\bar{u} - \bar{d})$ distribution. Eigenvector 14 is dominated by the gluon distribution, although the second largest contribution is from the down quark. Eigenvector 18 is entirely influenced by the $(s - \bar{s})$ distribution. It can be concluded that by using the LO PDFs, there can be limited improvements in the fit for certain distributions.

At NLO, however, the situation seems somewhat different. The POWHEG calculation is improved significantly by a few eigenvector sets, most notably 7, 5 and 14, which each contribute an improvement of approximately 0.05 per point in fit quality. The magnitude of the improvements are comparable to the LO case, but there are more eigenvectors which are contributing an improvement in this case. The NLO eigenvectors are different to those calculates at NLO, so it is not surprising that the influential sets are different to the Herwig++ calculation. The underlying distributions are, however the same. Eigenvector 7 and 14 are dominated entirely by the valence down quark, whereas 5 is mostly due to the $s - \bar{s}$. The relevant fractional contributions to the eigenvectors for both of the calculations are shown in Fig 2.11.



(a) LO - Eigenvector 2 (b) LO - Eigenvector 14 (c) LO - Eigenvector 18

(d) NLO - Eigenvector 7 (e) NLO - Eigenvector 5 (f) NLO - Eigenvector 14

Figure 2.11.: Fractional contribution to the uncertainty of an MSTW eigenvector due to a particular PDF. The chosen plots represent the dominant PDF contribution to the dominant eigenvectors in the Herwig++ fit (top row), and POWHEG fit (bottom row). Plots taken from MSTW paper [4].

In can be concluded that at next to leading order, the charged current inclusive jet cross sections may have some influence on the current knowledge of the underlying PDFs. The distributions most sensitive are the valence down quark and $(s - \bar{s})$. This is a notable result, since the only data currently used by MSTW which has an effect on the $(s - \bar{s})$ are the dimuon data sets. This data set could, then, be used to constrain the down and strange quarks, and help separate the valence down from the valence up distribution, a result that would have significant effects on measurements of the W mass at the Tevatron.

Chapter 3.

ATLAS & CMS Inclusive Jets

When considering hadron collider data for the determination of PDFs, one of the most effective and distinguishing sets is the inclusive jet cross section. The processes are almost entirely mediated by QCD matrix elements, and so the dynamics of the initial state partons can be easily probed. There are also electroweak corrections [65] which could potentially be reasonably large, however up to now there is disagreement upon the nature of these corrections, and so they are omitted from the analysis.

Until recently, the only hadron collider jet data which was available for PDF fits was that from the Tevatron measured by the CDF [66] and D0 [67] collaborations. These were shown to have a large effect on the PDFs due to the greater region of (x, Q^2) phase space probed. The introduction of LHC data is expected to have a similar impact on the current modern PDF sets.

The hadron collision jet data included in the MSTW 2008 PDF sets is in fact limited to only inclusive jet production, and the dijet data [68] was not used. This was due to large theoretical uncertainties from the scale definition, and an expected strong correlation with the better understood inclusive jet cross section. The inclusive samples were chosen due to a lack of theoretical understanding of the high rapidity dijet production as a function of dijet mass, M_{JJ}. This will be studied in more detail in the next chapter. In this chapter, the details of the theoretical prediction for inclusive jet cross sections at the LHC are studied and the effects they have on the PDFs are analysed.

This study was conducted after the publication, by the ATLAS collaboration, of the inclusive jet and dijet cross sections at 7 TeV using 36 pb^{-1} of data [52]. This was

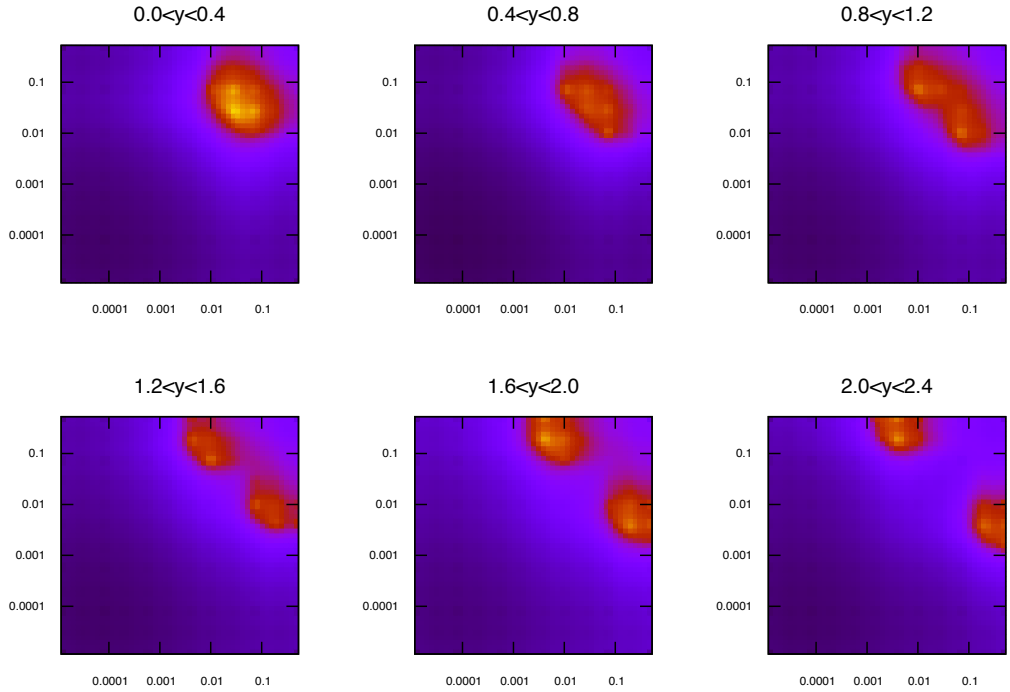
the first LHC data to have a true ability to probe new regions of the (x, Q^2) plane for current PDFs. To demonstrate this ability, Figure 3.1 shows the distribution of the parton momentum fractions x_1 and x_2 for NLOJet++ events at the Tevatron and the LHC. In the highest rapidity bin, the ATLAS data is probing values of $x \approx 10^{-5}$, 2 orders of magnitude lower than at D0. These plots are dominated by the low p_T bins within each rapidity bin, due to the orders of magnitude greater number of jets produced at low p_T. The higher p_T jets require higher x values, and the spots in Fig 3.1 shift along the diagonal line defined by $x_1 = x_2$ towards higher x as the p_T of the jets is increased. Comparing the plots at LHC and Tevatron energies shows the value of the LHC data.

3.1. Kinematics

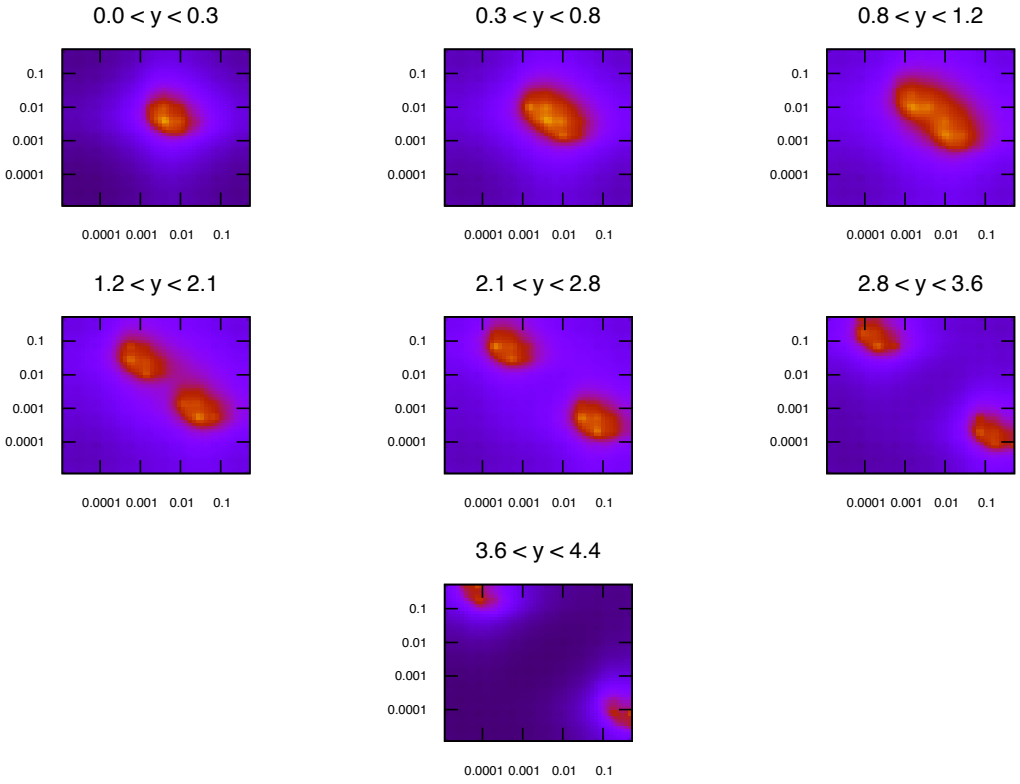
The sensitivity of the data to different partons is demonstrated in Fig 3.2, where the cross section calculation is broken down into its partonic subprocesses. Although there are 7 subprocesses within the calculation, these can be simplified to 4: gluon-gluon, quark-gluon, quark-quark, antiquark-antiquark.

Clearly, different areas of phase space provide more information about certain PDFs than others. In the lowest rapidity bin for instance, the low p_T jets are produced predominantly by initial state gluons, whereas the hardest jets are dominated by the quark-quark process. By combining this information with that obtained from Figure 3.1, we can see that the low-p_T central jets will provide information on the low x gluon, whereas high-p_T will shed light on the high x valence quark distributions. The fraction of partons probed also change as a function of rapidity. As the rapidity of the inclusive jets increases, the events are produced predominantly by a combination of one low x and one high x parton, which can again be seen in the plots of Fig. 3.1. This means that the quark-gluon process becomes dominant at high rapidities, especially at high p_T, and so these bins in the data will simultaneously probe the gluon and the quark distributions.

The χ^2 used to compare data to theory is similar to that used in MSTW PDF fits for jet data. Each data point is allowed to move with respect to the theory prediction due to the many systematic uncertainties in the measurement. For each source of systematic uncertainty, a nuisance parameter r_k is introduced, such that shifts will

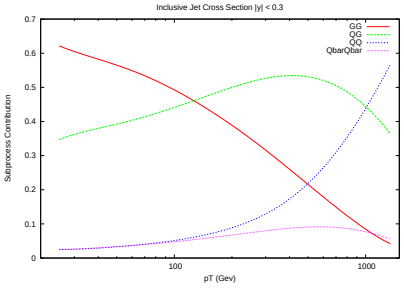


(a) D0

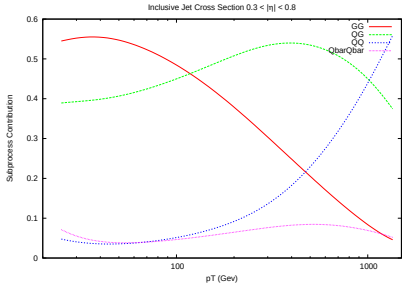


(b) ATLAS

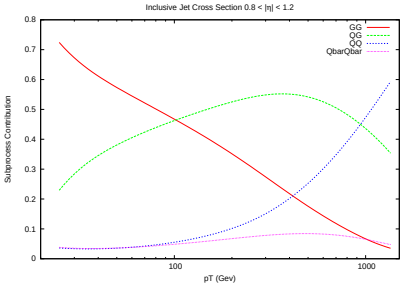
Figure 3.1.: Values of x_1 (highest x) and x_2 (lowest x) for each event generated in NLOJet++ for inclusive jets at the Tevatron ($\sqrt{s} = 1.96$TeV) and LHC ($\sqrt{s} = 7$TeV). The lowest p_T jets dominate in each rapidity bin, so the higher values of x probed at large p_T do not appear.



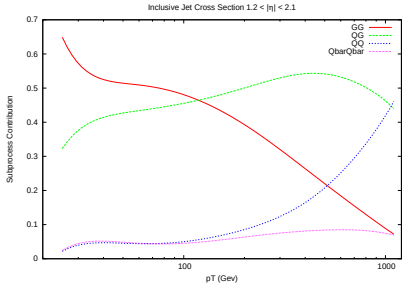
(a) $y < 0.3$



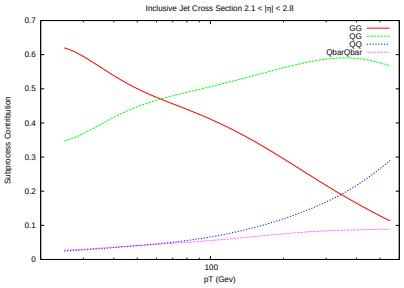
(b) $0.3 < y < 0.8$



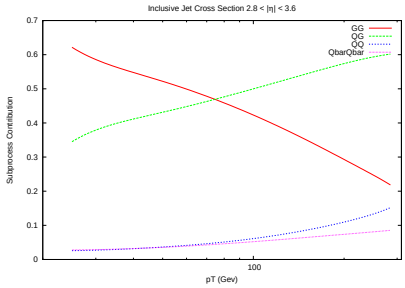
(c) $0.8 < y < 1.2$



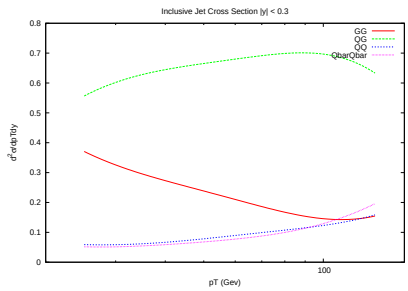
(d) $1.2 < y < 2.1$



(e) $2.1 < y < 2.8$



(f) $2.8 < y < 3.6$



(g) $3.6 < y < 4.4$

Figure 3.2.: Contributions of different initial-state parton combinations to the inclusive jet cross section calculation at ATLAS.

only occur if the reduction in χ^2 is significant. The exact form of the expression used is:

$$\chi^2 = \sum_{i=1}^{N_{pts}} \left(\frac{D_i - \sum_{k=1}^{N_{corr}} r_k \sigma_{k,i}^{corr} - T_i}{\sigma_i^{uncorr}} \right)^2 + \sum_{k=1}^{N_{corr}} r_k^2 \tag{3.1}$$

where i labels the individual data points and k labels the correlated systematics. In this case, the number of correlated systematics is 88 when including the hadronisation uncertainty. The uncorrelated error is the sum in quadrature of the statistical error and the 3 uncorrelated systematics. This definition is not identical to the standard MSTW fit due to the treatment of normalisations, which here is considered a standard source of systematic error. In the actual fits, the normalisations are treated seperately, and this will be discussed later in the chapter.

It is possible to solve this equation for r_k analytically, giving the optimum systematic shifts directly. By minimising the χ^2 the result is:

$$r_k = \sum_{k'=1}^{N_{corr}} (A^{-1})_{kk'} B_{k'} \tag{3.2}$$

where

$$A_{kk'} = \delta_{kk'} + \sum_{i=1}^{N_{pts}} \frac{\sigma_{k,i}^{corr} \sigma_{k',i}^{corr}}{(\sigma_i^{uncorr})^2}, B_k = \sum_{i=1}^{N_{pts}} \frac{\sigma_{k,i}^{corr} (D_i - T_i)}{(\sigma_i^{uncorr})^2}. \tag{3.3}$$

This means that by calculating and subsequently inverting the 88×88 matrix A, and the vector B, the optimal values of the nuisance parameters can be found.

The correlated systematics for both the inclusive and dijet data sets are mostly antisymmetric, and so a method of symmetrising to obtain a single error for each data point must be employed. Since this is a matter of choice and should not effect

the results in any meaningful way, three opposing methods were used to test the effect. These were:

$$\sigma_{corr} = |{\sigma_{corr}}^{+}|$$
$$\sigma_{corr} = |{\sigma_{corr}}^{-}|$$
$$\sigma_{corr} = \frac{(|{\sigma_{corr}}^{+}| + |{\sigma_{corr}}^{-}|)}{2}$$

where ${\sigma_{corr}}^{+/-}$ are the positive and negative values of the antisymmetric errors. The difference in χ^2 obtained from these methods varied by no more than 3% across all theory predictions. In the following results, the third definition is used to calculate the χ^2 values, due to some data bins displaying highly antisymmetric errors. With either the first or second choice, this could lead to either an overestimation or underestimation of the uncertainty. Whilst a small difference in the χ^2 fit is seen between the definitions, there is very little difference observed in the physical results from each definition.

3.2. PDF Reweighting

A useful tool to extract information on the partons affected by a data set is to analyse the change in fit quality when using the different eigenvector sets in a global PDF fit. The global minimum of the PDF set will not necessarily give the best fit to any individual data set, due to competing influences from other sets used in the global fit.

In order to see if there is any impact on the PDFs from a new data set, the reweighting procedure suggested in [70], [71], [72] and more specifically for MSTW PDFs in [73] is used (see also [74]). Firstly, the prediction for each eigenvector in the MSTW2008 fit is produced. These predictions are combined to produce 1000 PDFs randomly distributed in eigenvector space, using the formula:

$$F(S_k) = F(S_0) + \sum_{j=1}^{n} [F(S_j^{\pm}) - F(S_0)]|R_{jk}| \tag{3.4}$$

where R_{jk} is a Gaussian-distributed random number. The $F(S_k)$ can be any observable calculated using an eigenvector S_k, however in this treatment they are simply the PDFs themselves. By sampling the eigenvector sets directly and weighting each PDF equally, an accurate estimate of the Hessian error on each PDF is obtained without the need of referring to the eigenvectors of the covariance matrix. The central PDF is estimated simply by taking the average of these unweighted PDFs. Although this does not exactly reproduce the global minimum, the deviations are small and always well within the $1-\sigma$ error band. The source of these deviations is the nonlinear dependence of the parameters on x.

Each random PDF is weigthed according to its χ^2, and by statistical combination can provide an updated ideal PDF for the data set in question. The weighting formula is

$$w_i(\chi_i^2) = \frac{W_i(\chi_i^2)}{\frac{1}{N_{pdf}} \sum_{j=1}^{N_{pdf}} W_j(\chi_j^2)}, \quad W_i(\chi_i^2) = [\chi_i^2]^{\frac{m*(N_{pts}-1)}{2}} \exp\left(-\frac{\chi_i^2}{2}\right) \tag{3.5}$$

where χ_i^2 is the fit quality of the ith random PDF, N_{pdf} is the number of random PDFs generated and N_{pts} is the number of points in the fit. The weighting function is modified to include a multiplying factor m, to account for the case where the fit gives a χ^2 significantly better than 1 per point. In this instance, the weight function has a turning point, and so assigns lower weights to the best fits than those slightly worse. This is demonstrated in Fig 3.3, where all random PDFs give a better fit than 1 per point (for ATLAS inclusive jet data). In this case, a value of $m < 1$ is required to ensure the weights are assigned correctly. The actual value of m to choose will affect how quickly the weights decrease as the fit worsens, however the effect of this on the final reweighted PDFs will be shown to be negligible, and simply ensuring that the function does not turn over is sufficient.

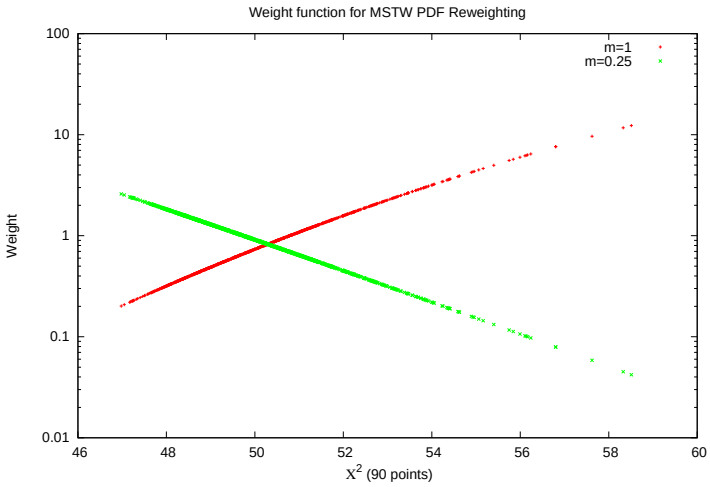


Figure 3.3.: Weights for 1000 random PDFs, each fit to a data set of 90 points with many PDFs giving χ^2 better than 1 per point. In this instance the standard reweighting function breaks down, and a value of $m < 1$ is needed to properly weight the PDFs.

A number which has can provide more information on the reweighting procedure is N_{eff}, the effective number of PDFs included in the reweighted distribution. This is calculated by:

$$N_{eff} = exp\left(\frac{1}{N_{pdf}}\sum_{i=1}^{N_{pdf}} w_i \ln\left(\frac{N_{pdf}}{w_i}\right)\right). \tag{3.6}$$

If the data set reweighted to has no effect, then all weights are 1 and $N_{eff} = N_{pdf}$, however as soon as there are some weights larger than others, N_{eff} will provide an estimate for the number of random PDFs which have contributed.

3.3. ATLAS Inclusive Jet Fit

Fig 3.4 shows the ratio of data to theory for the ATLAS 7 TeV R=0.4 inclusive jet cross section, both before and after the correlated systematics are taken into

account. The former gives a very poor agreement, with all data points above theory by up to 40%. The systematics are, however, large and the shifted points, defined as $(D_i - \sum_{k=1}^{N_{corr}} r_k \sigma_{k,i}^{corr})/T_i$ are almost all within 1σ of 1. The (R=0.4) data set is chosen over the R=0.6 due to the much smaller hadronisation corrections in the case of the smaller jet parameter.

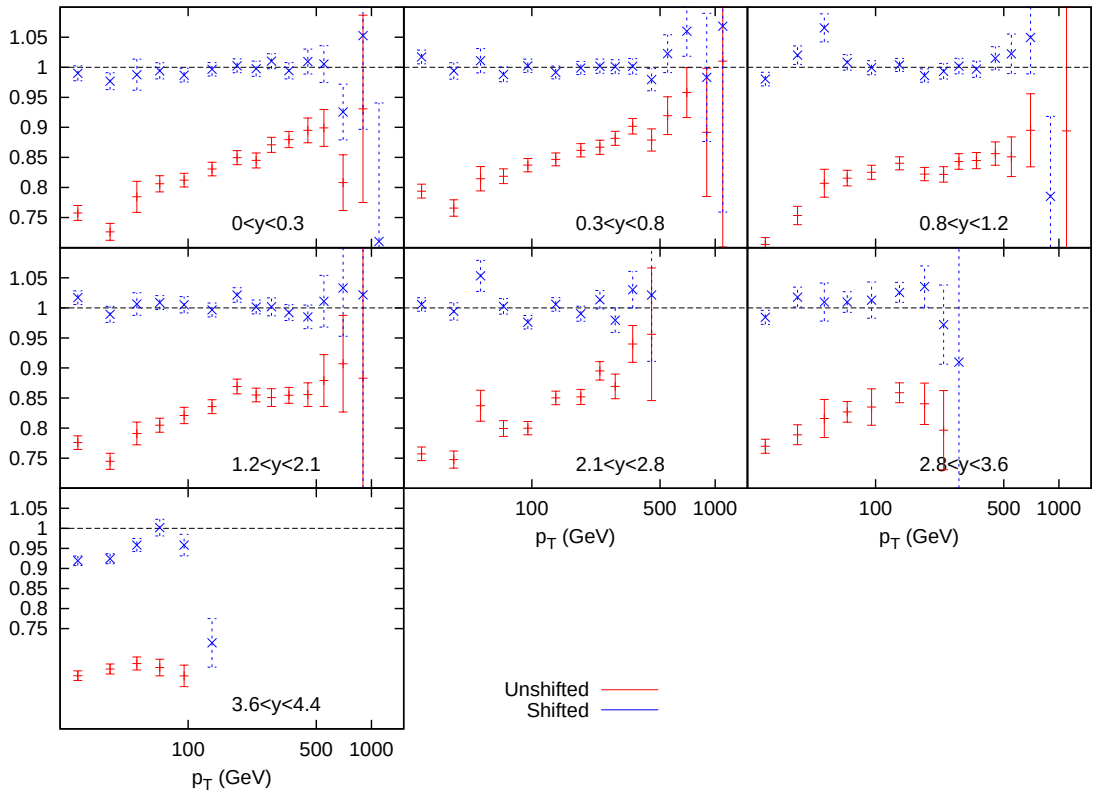


Figure 3.4.: Ratio of data to theory for ATLAS inclusive jets (R=0.4).

Scale	$p_T/2$	p_T	$2p_T$
R=0.4	0.75	0.78	0.70
R=0.6	0.85	0.79	0.72

Table 3.1.: χ^2 per point (90 points).

Tables 3.1 and 3.2 demonstrate that a χ^2 of less than 1 per point is achieved for all scale choices and both R parameter choices, whilst the vast majority of the r_k

$\lvert r_k \rvert <$	0.5	1.5	2.5	3.5
R=0.4	72	15	1	0
R=0.6	74	13	1	0

Table 3.2.: Distribution of r_ks (Total 88).

penalty terms are less than 0.5. This implies that the fit is a very good one, however the large shifts observed in the data alongside the small penalty terms implies that the systematic uncertainties are very large, and are drowning out any underlying physics effects.

The over compensation of systematic effects is further shown in Figs 3.5 and 3.6, where the effects of the data on the MSTW 2008 set is shown. Firstly, the individual eigenvectors are varied, and predictions produced corresponding to 1σ deviations in each direction. The change in χ^2 is negligible for all eigenvectors, with a maximum improvement of 0.007 per point in the R=0.4 fit for eigenvector 11. The results of the reweighting procedure are then shown in Fig 3.6 for the gluon, which is the only PDF noticeably affected by the data. There is a very slight trend for the gluon to increase at low x and decrease at high x, but again it is clear that very little can be deduced with the swamping effect of the systematics. The reweighted PDF receives a χ^2 of 0.73, from an unweighted value of 0.78. Again, this is testament to the small discerning power of the data set.

3.4. Additive vs. Multiplicative Errors

Another issue regarding the treatment of systematics is that of multiplicative or additive definitions. The errors are presented in the data as percentages, and so in order to obtain an absolute value of any given error, this percentage must be multiplied either by the data values or theory. If the percentage errors are multiplied by the data, they are considered additive since they are equivalent to an absolute error, whereas if they are multiplied by the theory they are considered multiplicative. By the nature of this particular fitting method, the data points themselves are significantly shifted in one direction by the systematics before the χ^2 is evaluated (in this case upwards, since the theory lies above data in general). Therefore, if the

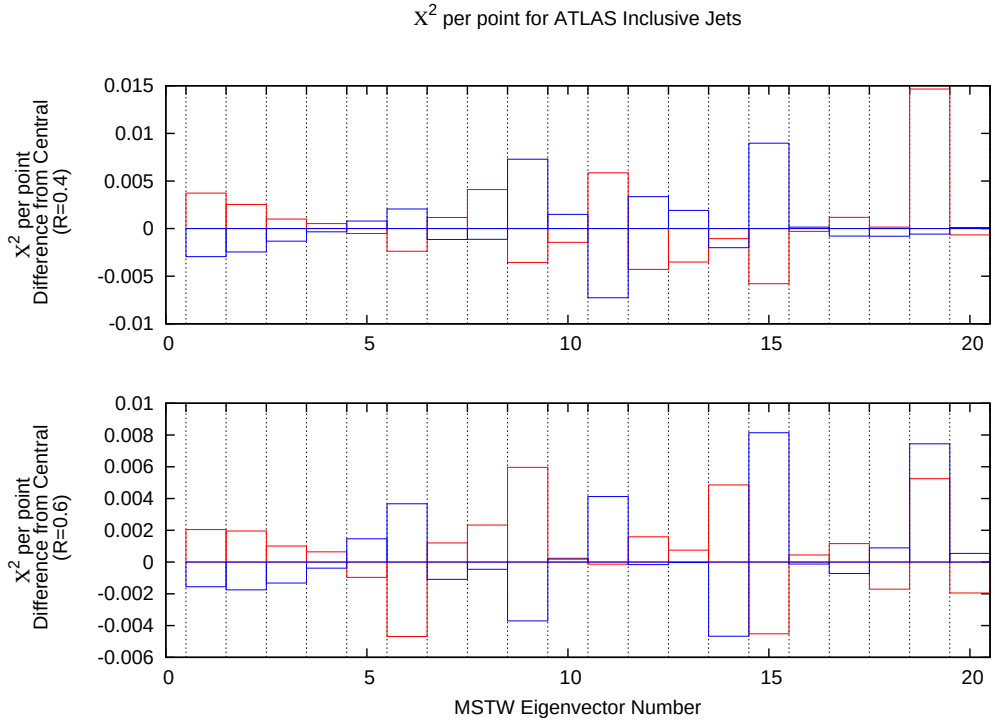


Figure 3.5.: Change in fit quality for each MSTW eigenvector direction for ATLAS inclusive jets for both R-parameters used. The blue (red) bars indicate positive (negative) movement in the eigenvector direction.

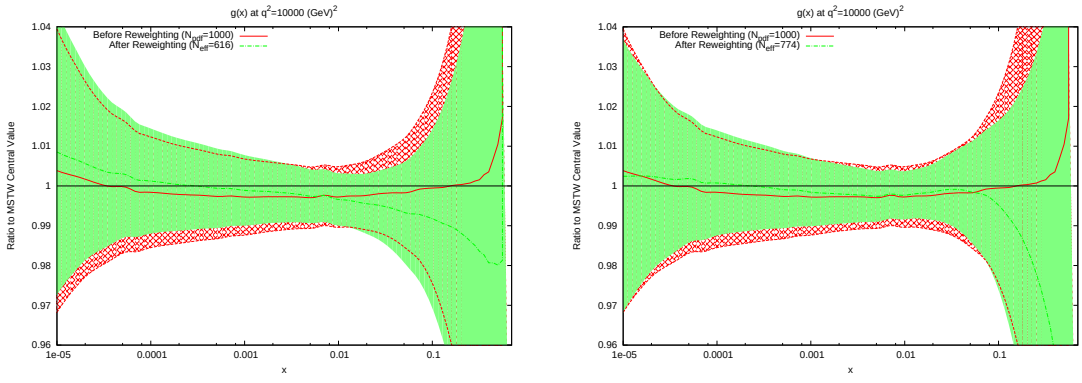


Figure 3.6.: The effect of reweighting the MSTW2008 gluon using ATLAS inclusive jet data. Jet size parameter $R = 0.4$ (left), and $R = 0.6$ (right).

absolute errors are obtained from the raw data, they will be proportionally smaller after the shift. The effect of this can be seen in Table 3.3 where the χ^2 for the two separate treatments of errors is summarised. The multiplicative treatment shows a considerably better fit than additive, due to the larger absolute size of each error.

Scale	$p_T/2$	p_T	$2p_T$
Multiplicative (R=0.4)	0.645	0.584	0.556
Multiplicative (R=0.6)	0.630	0.584	0.587
Additive (R=0.4)	0.752	0.773	0.703
Additve (R=0.6)	0.845	0.790	0.721

Table 3.3.: χ^2 per point using multiplicative and additive errors.

The table also demonstrates that the physics being probed depends upon the treatment of the errors. In the multiplicative case with R=0.6, the best fit is obtained with a scale choice of p_T, whereas it is $2 * p_T$ when using additive. Whilst it is a small discrepancy, it shows the importance of the treatment of errors, since everything else in the two fits is identical.

3.5. ATLAS Combined: 2.76 TeV and 7 TeV

A method for possibly reducing the effect of the systematic uncertainties of the inclusive jet cross section data is to perform a fit at two different centre of mass energies simultaneously, as done in [69]. The largest source of such uncertainties is the Jet Energy Scale (JES), which for ATLAS comprises of 14 separate uncertainties correlated across all bins in the measurement. Since the source of JES uncertainties is the same at any centre of mass energy, performing a PDF fit across two measurements will significantly reduce the allowed systematic shift of data points, allowing better constraints on PDFs.

The prediction for MSTW2008 is shown in Fig 3.7, both before and after the systematics shifts in the χ^2 calculation are taken into account. The data again must be moved upwards for all points in the combined set to match the theory, however when compared to the equivalent plot for the 7 TeV data (Fig 3.4), it can be seen

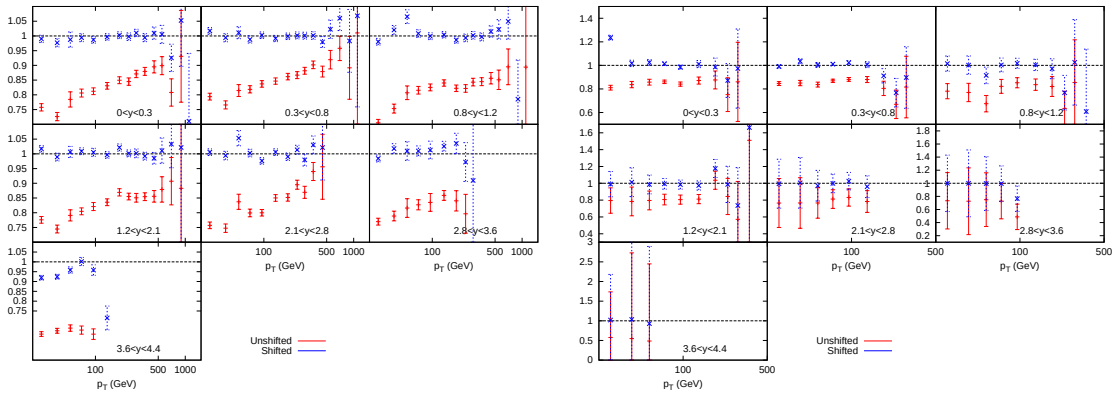


Figure 3.7.: Ratio of data over theory for MSTW PDFs convoluted with APPLgrid for the ATLAS inclusive jet combined data. The left hand plots are the 7 TeV data points, whilst the right hand side shows the 2.76 TeV data. There is more fluctuation in the shifted points for 7 TeV with the constraints imposed by concurrent 2.76 TeV fit, than for the pure 7 TeV fit.

that the systematics have less of an effect on this particular data set, with more fluctuations in the shifted points, especially at high rapidity.

Both the measurement of the inclusive jet cross section at 2.76 TeV and that at 7 TeV contain 21 sources of correlated systematic uncertainty which translate into 88 individual uncertainties after considering the correlations between rapidity bins. Only 3 of the sources are not correlated between the two data sets, and so the combined measurement contains 91 separate correlated uncertainties, an increase of only 3 whilst increasing the data points from 90 to 149.

3.5.1. Data Cuts

The original paper [69] to produce such a PDF analysis was produced by the HERAPDF collaboration in conjunction with ATLAS. In this analysis, a minimum p_T cut is applied of 45 GeV for all bins in both data sets, whilst the 2.76 TeV data set includes a further maximum p_T cut of 400 GeV applied in all but the $1.2 < y < 2.1$ bin. These cuts are motivated by the large hadronisation corrections in the stated bins, which can be as high as 12% for some low p_T bins. The hadronisation corrections for all other bins is at or below 2%. For this analysis, both definitions will be tested. The difference in fit quality for each of the major PDFs is shown in Table 3.4, where

a large improvement is seen when including the p_T cuts for various PDFs. The source of this improvement is from the low-p_T bins, where the statistical errors are the smallest, and so any deviation from the data (which will occur due to the large hadronisation uncertainties) produces a comparitively large increase in χ^2. The cuts are clearly important, since the ordering of the best PDF set changes for the two different treatments.

	No Cuts	HERAPDF Cuts	Additive Errors
MSTW 2008	1.43	0.91	1.37
NNPDF 2.2	1.42	0.89	1.43
HERAPDF 1.5	1.36	1.08	1.55
CT10	1.59	0.97	1.95
ABM11	1.41	1.10	1.94

Table 3.4.: χ^2 per point for ATLAS combined data, both with and without p_T cuts. The third column uses additive errors and has the two discussed anomalous points cut. NLO PDF sets are used.

The source of the increase in fit quality for the cut data can usually be traced to one or two points in the set. For example, The lowest p_T bin in the $0 < y < 0.3$ bin of the 2.76 TeV data set contributes over 100 points to the total χ^2 when using the MSTW2008 data set. However, when using NNPDF 2.2, there are two points which contribute 39 and 30 points respectively, and correspond to the lowest two p_T bins of the $3.6 < y < 4.4$ rapidity bin in the 7 TeV data. The points which contribute disproportionately large amounts to the fit are always located in the points which are cut from the HERAPDF analysis.

As discussed for the pure 7 TeV fit, the way in which the systematic errors are treated is important to the quality of fit due to the systematic shift between data and theory. In the previous section the multiplicative definition has been used since this is the treatment which most closely follows the HERAPDF/ATLAS analysis. Now, the additive definition is discussed. Since the same shift upwards from the data to the theory is seen in the ATLAS combined data set, it is expected to give a worse fit. This is true, and for MSTW2008 NLO PDFs, the fit becomes 2.44 per point, more than doubling the χ^2 from the multiplicative treatment. However, the majority of this χ^2 is localised to two anomalous points, even after the HERAPDF cuts, which

contribute approximately 130 points to the total χ^2. These are the highest p_T bin of the highest rapidity bin of the 7 TeV data, and the lowest p_T bin (after cuts) of the third rapidity bin of the 2.76 TeV data. Removing just these two additional points reduces the χ^2 to 1.37 per point. Since the MSTW fitting code uses additive errors for all data sets, it was proposed to remove these points for a PDF fit including this data. The effect on the PDFs was studied both with and without these two points, and no significant difference was found. The χ^2 values for each major PDF group is shown in the third column of Table 3.4.

3.5.2. NNLO PDFs

When considering NNLO PDFs, it is necessary to use NNLO matrix elements for the theoretical predictions. For hadron-hadron inclusive jet cross sections, these calculations have to date not been produced, and so approximations must be utilised to obtain the theoretical cross sections. The approximation used in the MSTW2008 analysis for Tevatron inclusive jets is based on the calculation by Kidonakis and Owens [75]. This calculation produces a threshold resummation which is based around the assumption that the parton-parton scattering phase space is restricted to the threshold region of $x_T = 2p_T/\sqrt{s} \sim 1$, due to the rapid decrease in PDFs at high x. The corrections are provided within the FastNLO framework, and so have been included for the use of Tevatron inclusive jet data in NNLO fits.

In order to include the LHC data into an NNLO fit, the threshold corrections are now calculated for the new data. The results are shown in Fig 3.8, where the ATLAS data is presented alongside that of D0. The first point of note is that the LHC phase space spans a region which extends much further from the threshold region than the Tevatron. The Tevatron threshold corrections maintain a sensible correction of approximately $\sigma_{NNLO} \sim 1.1\sigma_{NLO}$ across the majority of the phase space, however this correction clearly increases away from threshold. The corresponding ATLAS calculation demonstrates that this trend continues even further, and although the central jets maintain a reasonable correction throughout, the forward jet corrections become very large with decreasing x_T.

It is clear that for LHC jets, it will be necessary to include the full NNLO matrix elements in order to perform a full NNLO fit. Although not yet fully performed, the gluon-gluon process has been calculated by Gehrmann, de Ridder et al. [76]. These

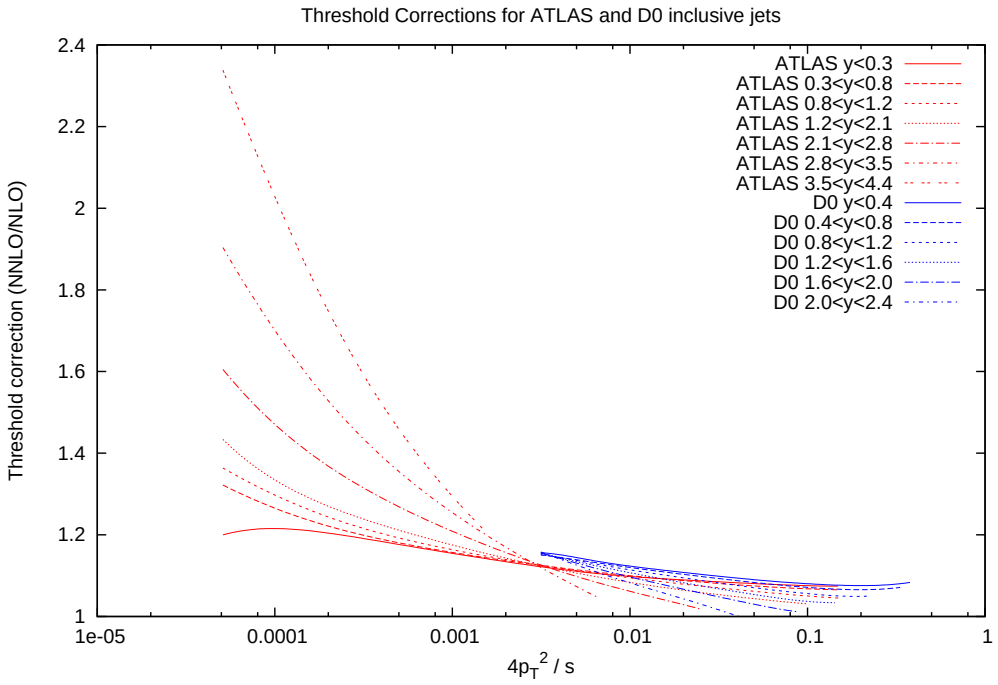


Figure 3.8.: Comparison of NNLO threshold corrections for ATLAS and D0 inclusive jets as a function of $x_T^2 = 4p_T^2/s$.

calculations have shown a correction of between 1.1 and 1.3 of the NLO prediction across all jet p_T values, and suggest that the threshold corrections indeed are not applicable to the LHC scenarios, although for jets at the Tevatron the corrections seem reasonable. As a result, the NNLO PDF predictions for the ATLAS jet cross sections will here be studied using the NLO calculation, with the note that any future NNLO fit which includes such data should use the full NNLO calculation when released.

The fit values for each of the major PDFs at NNLO are shown in Table 3.5. All of the major PDF sets with the exception of ABM11 do well in describing the data after the low p_T cuts are applied. This can be seen in the relative shape of the gluon at NNLO for each of the groups, shown in Fig 3.9. Whilst there is in general disagreement between all of the PDFs at NLO, the NNLO sets show a common trend for the four which provide similar good fits, with ABM11 showing a significantly

higher gluon across most x values, and large disagreement with the other PDFs at high x, although here the PDF errors are very large. Again the additive error treatment with the additional cuts is shown in the third column. The ordering changes with HERAPDF providing the best fit, and all of the PDFs provide a good fit with this treatment.

	No Cuts	HERAPDF Cuts	Additive Errors
MSTW 2008	1.32	0.927	1.44
NNPDF 2.2	1.41	0.907	1.31
HERAPDF 1.5	1.36	0.921	1.26
CT10	1.47	0.907	1.42
ABM11	1.97	1.76	1.64

Table 3.5.: χ^2 per point for ATLAS combined data, both with and without p_T cuts. The third column uses additive errors and has two additional anomalous points cut. NNLO PDF sets are used.

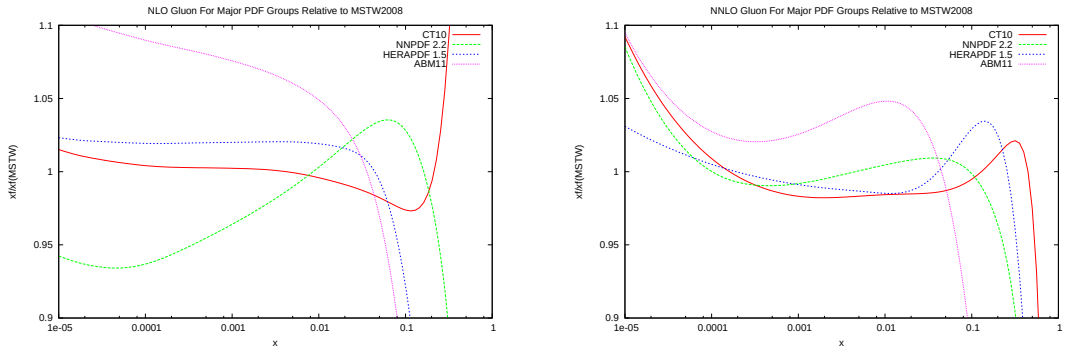


Figure 3.9.: Comparison of the major PDF groups at NLO and NNLO.

3.5.3. Effect on MSTW PDFs

The effect on the PDFs using the reweighting technique is shown for the case of multiplicative errors in Fig 3.10 and for additive errors in Fig 3.11. In both cases, the central value of the reweighted gluon is consistent with standard MSTW 2008 central value across all values of x and in the multiplicative case it is very similar to the

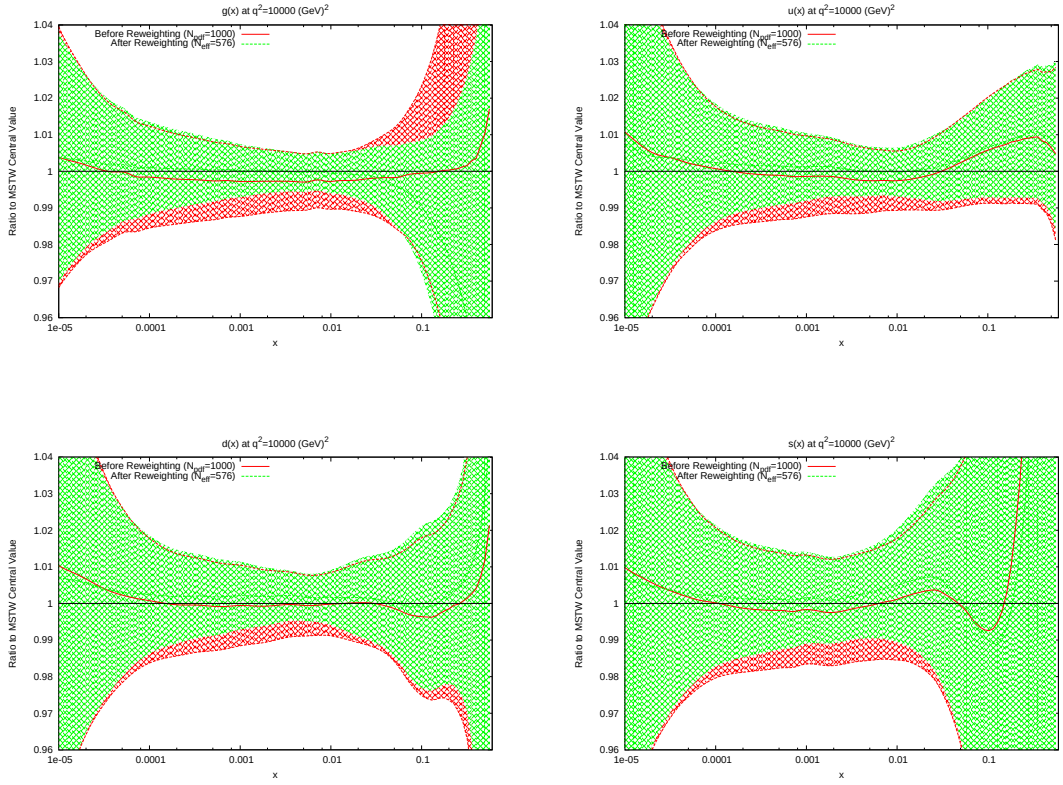


Figure 3.10.: Effect of the ATLAS combined inclusive jet data on the gluon and quark PDFs. Here, multiplicative errors are used, and the lowest two bins in p_T in all rapidity bins and the highest p_T bins in the 2.76 TeV rapidity bins are excluded as per the HERAPDF analysis.

reweighted pure 7 TeV gluon. The error bands are reduced in size more significantly than when using just the 7 TeV data, and the additive treatment seems to have more of an effect in this sense than the multiplicative. The upward shift in the quark PDFs and also the error constraints are larger when using the additive treatment. Clearly there is more constraint on the gluon with the 2.76 TeV data included, and the reduction in systematics is allowing more information on the PDFs to be extracted. The quark PDFs are also shown; although the effect is again larger than the pure 7 TeV case, there is very little movement from the central MSTW value. Another hint to the improvement in PDF extraction from using only 7 TeV is the reduction in the number of effective PDFs, N_{eff}, in the reweighting procedure for both of the error treatments. When using multiplicative errors, the χ^2 is reduced from 0.974 to 0.962 by reweighting, and for additive errors the effect is larger as expected from

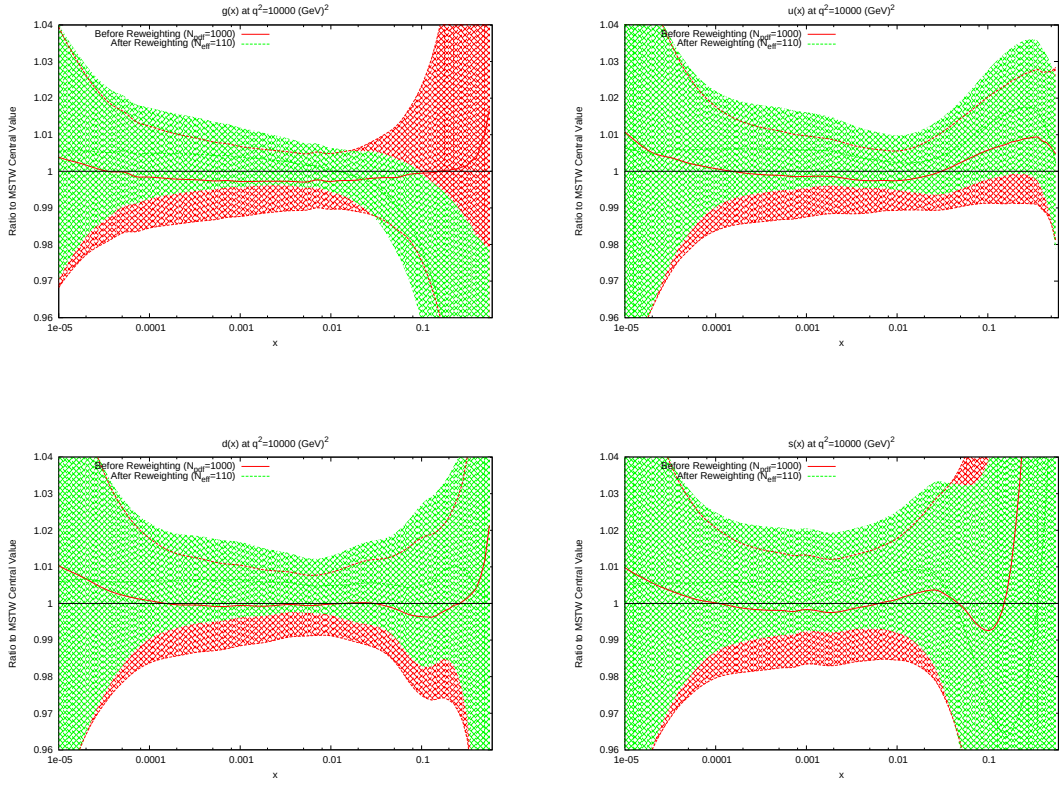


Figure 3.11.: Effect of the ATLAS combined inclusive jet data on the gluon and quark PDFs. Here, additive errors are used in the determination of the χ^2, and the two anomalous points discussed in the previous section are additionally cut.

the reweighted plots, changing from 1.45 to 1.26. Whilst the two additional data points cut were deemed necessary to provide a sensible fit, it was observed that even by inluding these points, the reweighed PDFs looked very similar for the additive treatment. Hence, the difference between Figs 3.11 and 3.10 can be attributed to the differing error treatments.

3.6. CMS Inclusive Jets

To date the LHC data set with the most resolving power for PDFs is that released by the CMS collaboration [53] in early 2013. This analysis, like the earlier ATLAS

analysis, was performed at 7 TeV. However, a much higher collected luminosity of 5 fb^{-1} is included, and so statistical errors are greatly reduced.

Compared to the ATLAS measurement, the jet p_T spectrum extends much higher to 2 TeV, however the minimum value is only 114 GeV. There is also less rapidity span for the CMS jets, which are only measured to 2.5. The overall effect is to have more pronounced sensitivity to high x PDFs, and lower sensitivity to low x PDFs. This can be seen in Fig 3.12, where the (x_1, x_2) distribution for each event generated is shown. The reach to low x is limited to 10^{-3}, but each distribution is shifted more towards the high (x_1, x_2) region.

The partons which are probed by the data are therefore naturally different from those of ATLAS. The greater emphasis on medium to high x partons means a greater relative contribution from quarks. Fig 3.13 shows the partonic composition of the calculation at each point in phase space. Unlike the ATLAS jets, the gg subprocess does not dominate anywhere in the phase space, with gq contributing maximally everywhere except for the very highest p_T jets.

The raw calculation using NLOjet++ interfaced with APPLgrid is in much better agreement with data than the ATLAS inclusive cross section. Whilst the ATLAS jet calculation was up to 30% too high in some bins, the CMS calculation is never more than 10% off. The systematics must again be taken account of in a χ^2 fit, and the comparison to data again improves after this consideration. However, as Fig 3.14 shows, the shifted data/theory points reflect the statistical fluctuations present in the unshifted points. For the ATLAS fit, it was clear that the statistical fluctuations were being washed out by the large freedom provided by the systematics.

Scale	$p_T/2$	p_T	$2p_T$
MSTW 2008	1.92	1.48	1.12
NNPDF 2.2	1.56	1.43	1.56
HERAPDF 1.5	2.80	2.61	2.39
CT10	2.10	1.83	1.57
ABKM09	1.94	1.86	1.80

Table 3.6.: χ^2 per point (133 points) for NLO PDFs for CMS inclusive jet data.

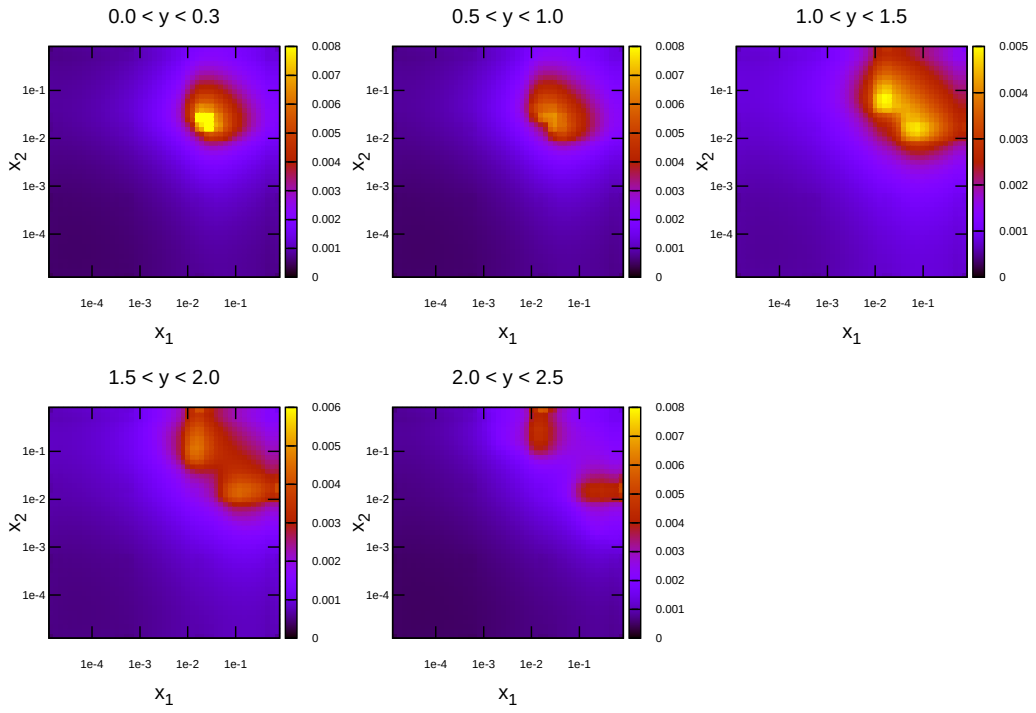
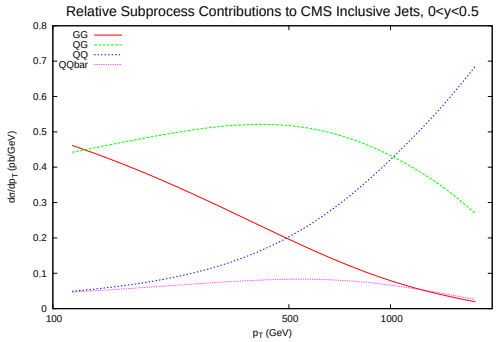
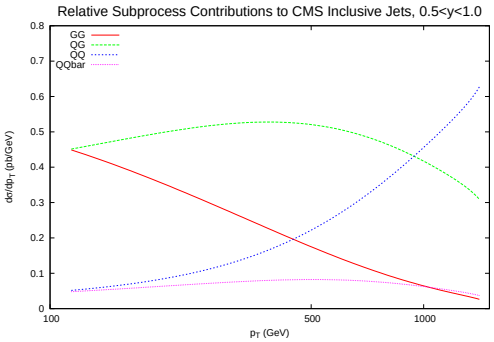


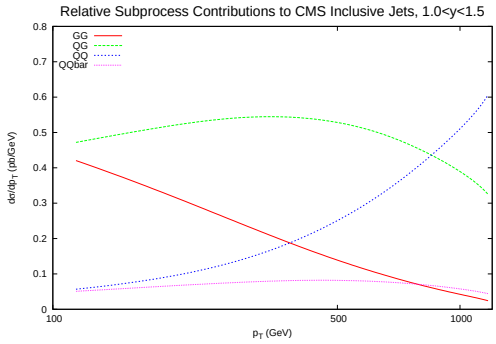
Figure 3.12.: Distribution of $x_{1,2}$ values for NLOjet++ events in the CMS inclusive jet calculation.



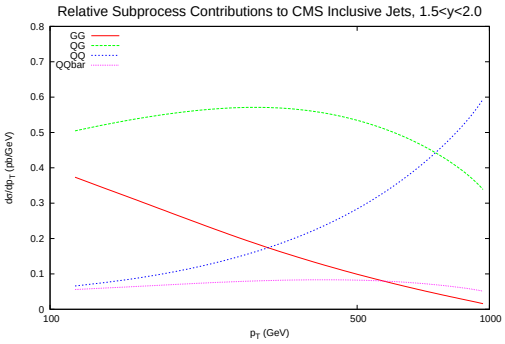
(a) $y < 0.5$



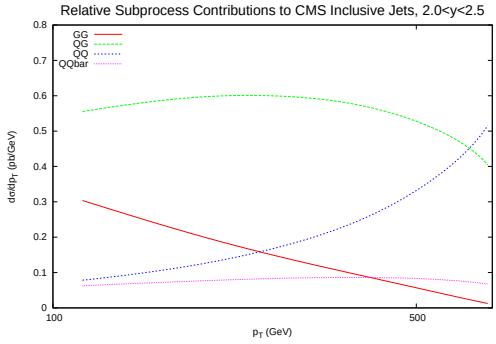
(b) $0.5 < y < 1.0$



(c) $1.0 < y < 2.5$



(d) $1.5 < y < 2.0$



(e) $2.0 < y < 2.5$

Figure 3.13.: Contributions of different initial-state parton combinations to the CMS inclusive jet cross section calculation.

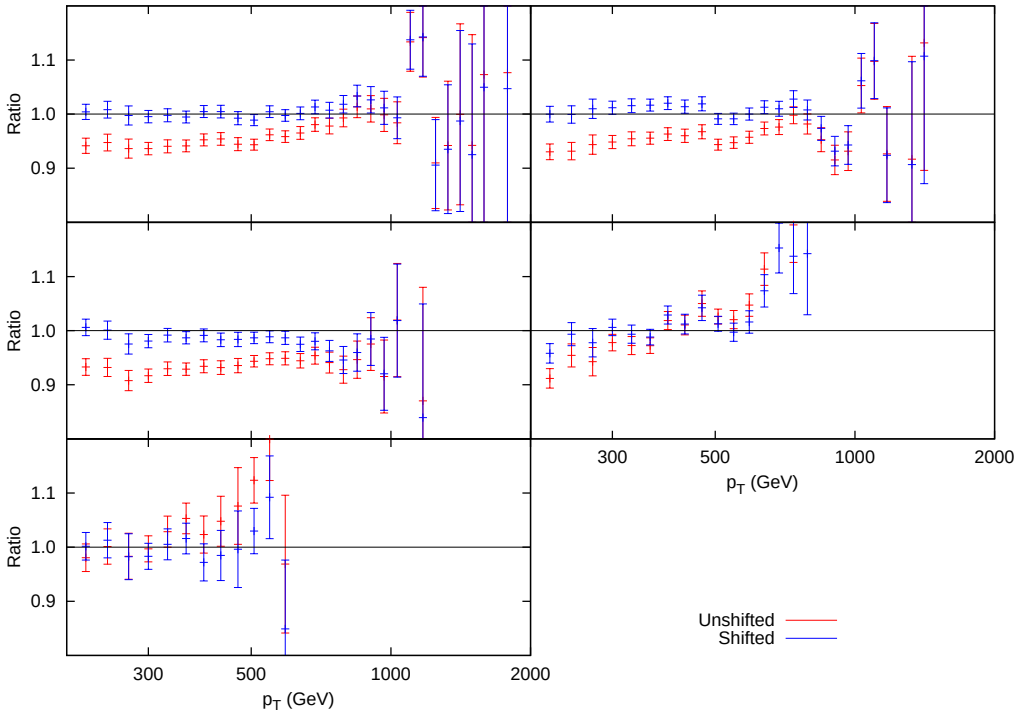


Figure 3.14.: Ratio of data to theory using MSTW 2008 NLO for CMS inclusive jets. Both the raw APPLgrid calculation and the calculation after systematic effects are taken into account are shown.

The table of fits is shown in Table 3.6, and the corresponding systematic shifts in Table 3.7. The χ^2 values are generally worse than for the ATLAS data, with HERAPDF 1.5 and ABKM09 noticibly poor. With fewer r_k values, it is clear that there is less freedom to compensate for differences by using the systematic shifts. This is reflected by the distribution of r_ks, which for ATLAS produced a majority below 0.5, but for CMS is skewed towards higher values. For both HERAPDF and ABKM, the luminosity r_k is required to be over 3.5 to obtain the best fit. If the luminosity is held fixed, both of these fits become worse by ~ 0.2 per point.

The predictions from each PDF group as a ratio to MSTW 2008 is shown in Fig 3.15. The central values of NNPDF and CT10 are close to the MSTW central value for most of the phasespace, and this is reflected in the similar χ^2 fits. HERAPDF and ABKM tend to lie below, and with similar shapes. The large luminosity r_k shift

$\lvert r_k \rvert <$	0.5	1.5	2.5	3.5	4.5
MSTW 2008	8	8	2	1	0
NNPDF 2.2	8	9	2	0	0
HERAPDF 1.5	6	9	1	2	1
CT10	8	7	3	1	0
ABKM09	8	7	3	0	1

Table 3.7.: Distribution of r_ks (Total 19).

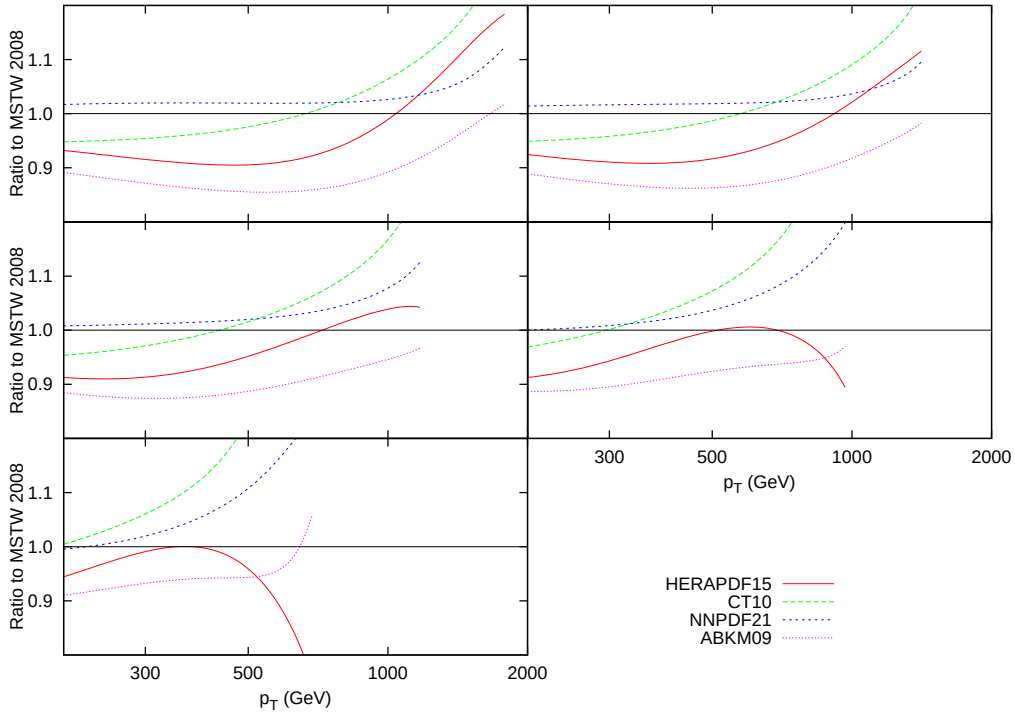


Figure 3.15.: The ratio of the main PDF groups' predictions to the MSTW 2008 prediction, for CMS inclusive jet production.

for these two PDFs can be understood through this plot, especially for ABKM which lies approximately 10% below the MSTW prediction for all data points.

As with the ATLAS inclusive jets, the NNLO PDFs are studied using NLO matrix elements. A similar trend is seen as in the previous case; the originally poor HERAPDF NLO fit comes more into line with the other sets, whilst ABM worsens from its NLO fit. The luminosity shift for HERAPDF is now a more sensible -1.5,

whilst for the ABM NNLO PDF it is 6.88. By fixing the luminosity, the fit worsens even further to 3.97 per point, and is clearly incompatible with the data.

Scale	p_T
MSTW 2008	1.37
NNPDF 2.2	1.39
HERAPDF 1.5	1.82
CT10	1.49
ABKM09	3.37

Table 3.8.: χ^2 per point (133 points) for NNLO PDFs for CMS inclusive jets.

3.6.1. Effect on MSTW PDFs

The same procedure as described for the ATLAS jets is applied to the CMS data set. The variations of the fit under movements in the eigenvector directions are shown in Fig 3.16. This time, there are significant improvements in some directions, with eigenvectors 8, 10 and 19 reducing the χ^2 the most. These eigenvectors are most influenced by the u_v, d_v and gluon distributions respectively.

When the reweighting procedure is applied, the results of which are shown in Fig 3.17, the effect is larger than the full ATLAS combined data set. The shape of the reweighted gluon agrees with the ATLAS reweighting, with a much lower gluon at high x. What is significant is the increased sensitivity to the quark PDFs. The reduction in error band in the up and down distributions is similar to that for the gluon. Even the error in the strange distribution is reduced in both directions across almost all values of x. The focus of the CMS data on higher values of x has lead to a less dramatic effect on the gluon, but consistently better constraining of all quark PDFs. The reweighting improves the fit quality from 1.47 to 1.29.

3.7. Summary

The inclusive jet data which has been measured during the first run of the LHC at 7 TeV is our first look at QCD in a new energy regime, and so testing our current

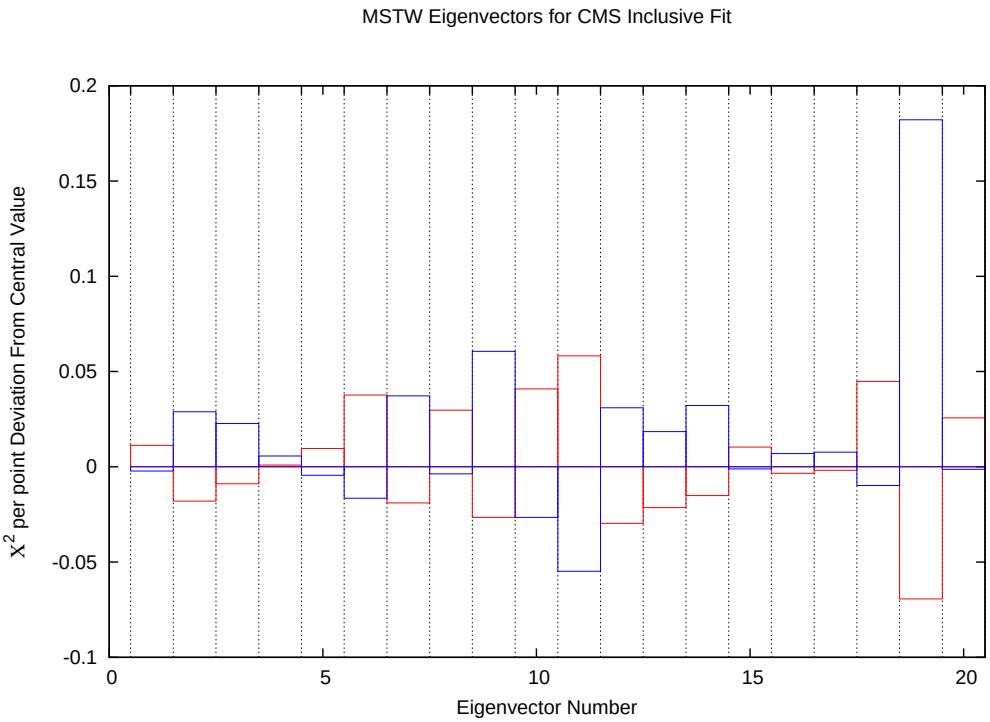


Figure 3.16.: Change in fit quality from the central MSTW2008 PDF for each eigenvector in the set.

knowledge of PDFs with this data is a vital task. The conclusion from these early runs is that the MSTW08 PDFs hold up well in this regime, since none of the data from either experiment has required a PDF to move outside its 1σ error band.

The earliest released measurement was the least discerning for PDFs; the ATLAS inclusive jet cross section at 7 TeV using 36 pb^{-1} of luminosity was inevitably dominated by systematic uncertainties, and so any variation in physics parameters used in this chapter are incapable of improving a fit in any way.

This issue is overcome by the inclusion of a simultaneous measurement at centre of mass energy 2.76 TeV. The cancellation of systematic effects associated with jet energy scale provides a more suitable environment for testing PDFs. In this measurement, too, a good fit is found for MSTW 2008 PDFs. Although the data prefers a larger low-x and softer high-x gluon, these movements are still entirely within the error bands. A significant improvement in error is seen for the gluon

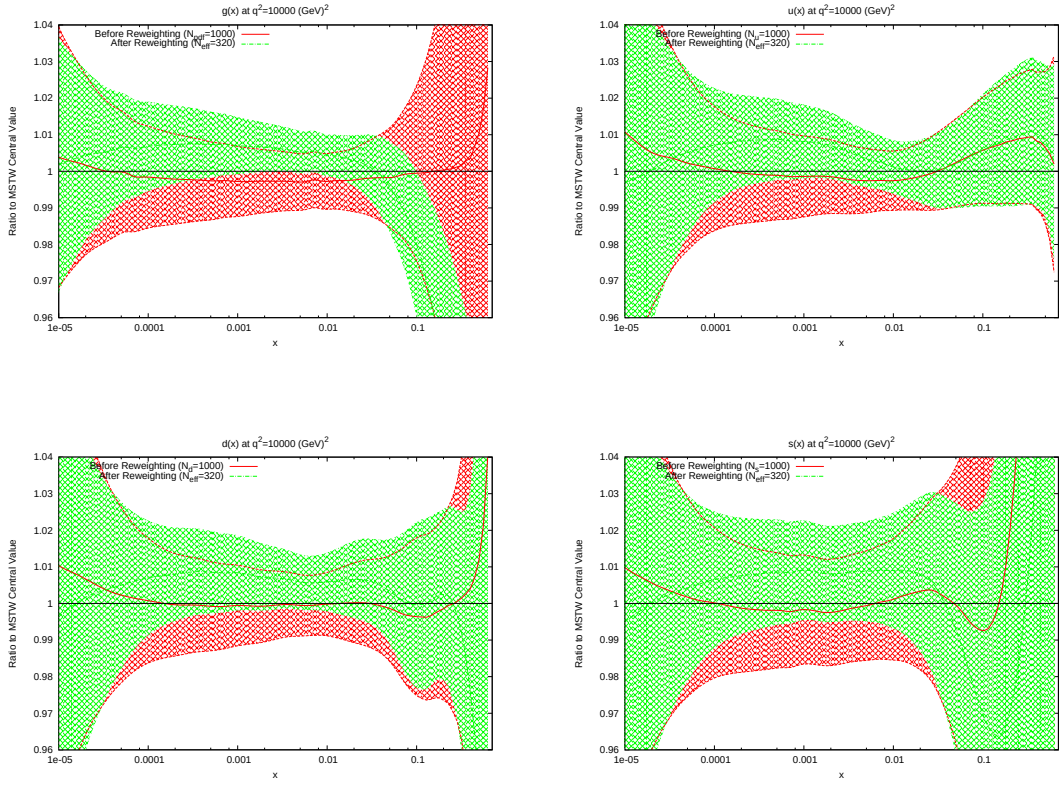


Figure 3.17.: Effect of the CMS inclusive jet data on the gluon and quark PDFs.

across all x, which implies that, if included in a new fit, this data could provide more accurate PDFs for the LHC era.

Finally the latest CMS data is analysed. With much higher luminosity than the ATLAS data, this is currently the measurement with the most potential for constraining the PDFs. Again a reasonable fit is found for MSTW, although the χ^2 per point is higher than the ATLAS fit, and some other PDFs do not appear to be compatible with the data. Due to the kinematics of the measurement, more focus is given to the quark densities for this set, and a reduction in the error bands is seen for all flavours. Again, including this data into a new fit would provide better constrained PDFs.

Chapter 4.

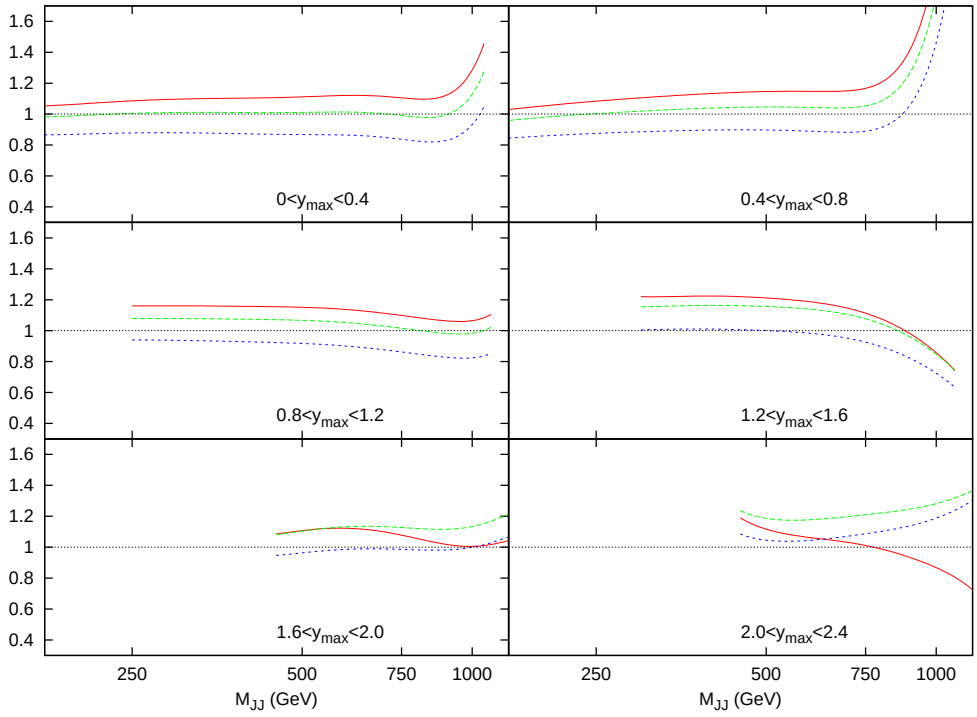
Dijet Cross Sections

In all previous MSTW fits, only inclusive jet data has been included into the fit. This is mainly due to uncertainties in the calculation of dijet cross sections and the scale choices therein. Whilst there is very limited scope for changing the kinematic choice of scale for inclusive jets, there are many possibilities when considering dijet cross sections. As such, this chapter presents a thorough study of the effect of the choice of renormalisation and factorisation scale choice on dijet predictions at both the Tevatron and the LHC, and the feasibility of including these datasets in a PDF fit is tested.

Before 2011, the only dijet cross sections available was from D0 at the Tevatron. Studies into the comparison were conducted [68], but inconsistencies in the scale uncertainty were found. The NLO calculations were performed using the average jet p_T as the scale choice, and this was shown to exhibit strange behaviour at high rapidities. This is demonstrated in Fig 4.1, where the predictions for 0.5, 1 and 2 times the scale choice are shown to cross over at high y_{max} and mass. In order to understand the source of this behaviour, the kinematics of the process must be studied.

4.1. Kinematics of Dijet Production

The kinematics of the dijet production process are defined using the invariant mass of the dijet system, M_{JJ}, and the rapidity of each of the jets in the event. A double-differential cross section is constructed using bins in the dijet mass and a combination



(a) p_T^{av}

Figure 4.1.: Theory/Data ratio for D0 dijets, using multiples of p_T^{av} as the choice of μ_R and μ_F. The multiples are 0.5 (red), 1.0 (green) & 2.0 (blue).

of the two rapidities. The flexibility in the latter leads to different possibilities for rapidity binning, and the D0 and ATLAS measurements use differing definitions. Where D0 uses y_{max}, the maximum rapidity of the two jets comprising the dijet pair, ATLAS chooses $y* = (y_1 - y_2)/2$, the difference between them.

This is the cause of the greatly differing x distributions of Figs 4.2 & 4.3. Using the maximum jet rapidity results in a similar pattern to inclusive jets, due to the fact that only the rapidity of one jet is considered. At high rapidities a single high-x parton must combine with a single low x parton, and low rapidities require equal values of x in both partons.

Using the rapidity difference, however, allows a much wider range of parton momentum fractions to produce dijets in all $y*$ bins. The observed shift towards high x at high $y*$ is in fact due to the fact that only high M_{JJ} events are measured at these rapidities. These high M_{JJ} events are also present in the other rapidity bins,

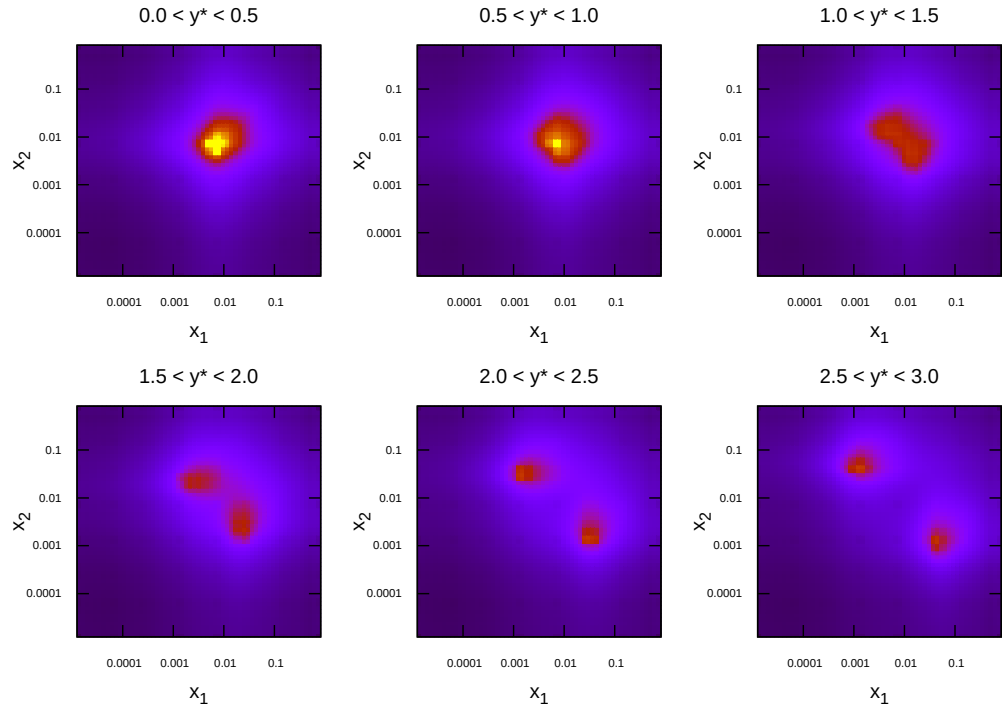


Figure 4.2.: Values of x_1 (highest x) and x_2 (lowest x) for each event generated in NLOJet++ for inclusive jets at the Tevatron ($\sqrt{s} = 1.96$ TeV) and LHC ($\sqrt{s} = 7$ TeV).

however due to the powerlike drop in cross section with dijet mass these events do not register in the respective plots and only the lowest mass bins can be seen.

The difference in the distributions of parton momenta leads to the question of which partons are being probed at different points in the phase space. Here we can begin to see the differences in the various datasets, especially when comparing to the relevant commensurate inclusive jet data. For the D0 dijet cross section in Fig 4.4, it is clear that the quark PDFs are in general the most important, with the gg luminosity always below the qg, and mostly below the $q\bar{q}$ luminosities.

For the ATLAS dijets, Fig 4.5 shows that for low rapidities, a similar behaviour to the corresponding inclusive jet plot is seen, with the gluon density dominating until the very high p_T (high M_{JJ}) region. However, at higher rapidities, the requirement of two high-x partons means the qq luminosity becomes by far the most important.

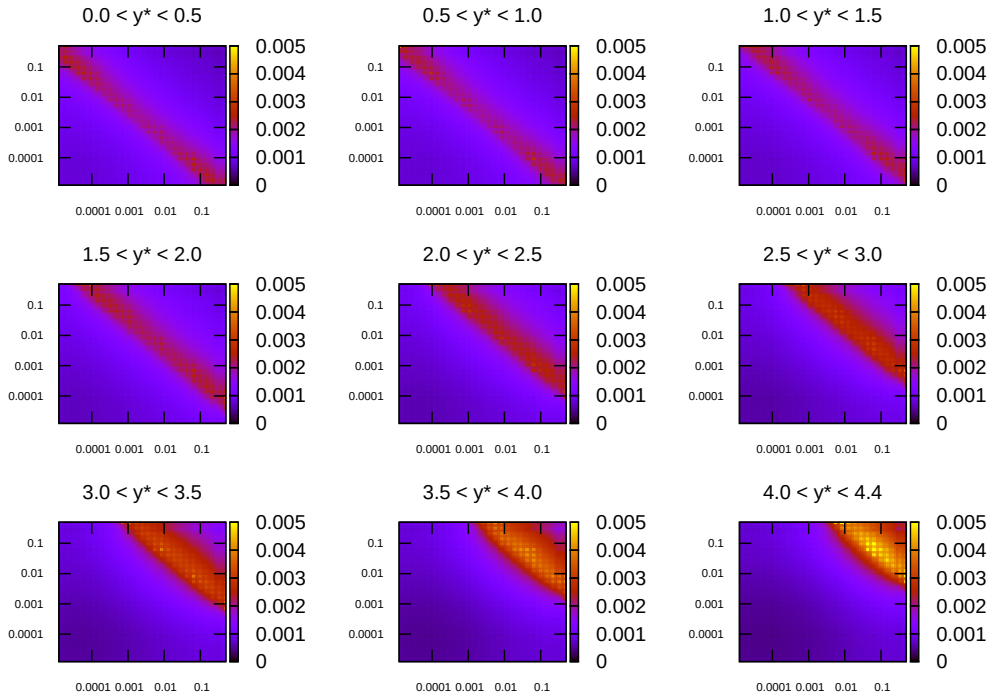
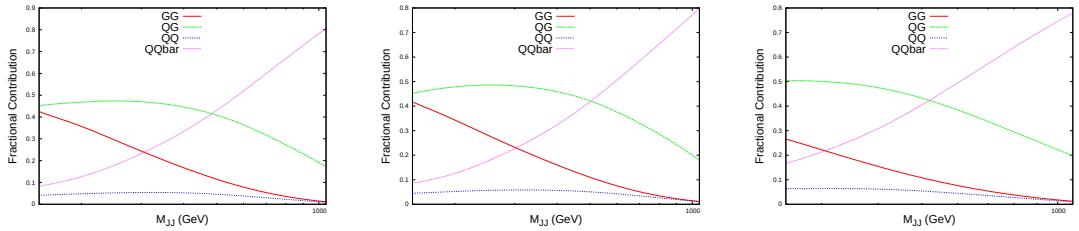


Figure 4.3.: Values of x_1 (highest x) and x_2 (lowest x) for each event generated in NLOJet++ for dijets at the Tevatron ($\sqrt{s} = 1.96$ TeV) and LHC ($\sqrt{s} = 7$ TeV)

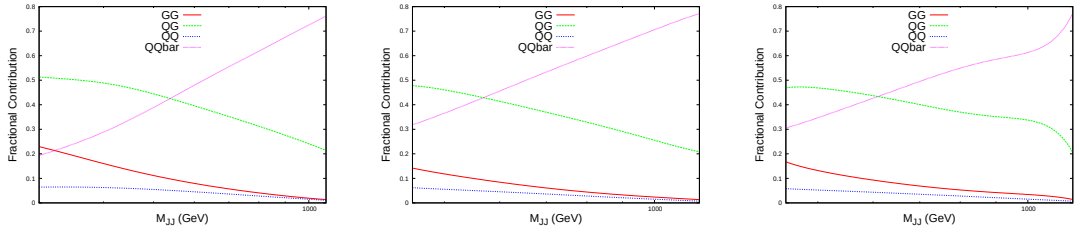
As a result, the dijet data for ATLAS should effect the quark densities far more than when using only the inclusive data.

4.2. Scale Variations

When considering dijet production, the choice of renormalisation and factorisation scales to include in the NLO calculation is not obvious. In general the behaviour of varying the scale on the full NLO calculation performed by NLOjet++ can be seen in the form of the differential cross section:



(a) $y < 0.5$ (b) $0.5 < y < 1.0$ (c) $1.0 < y < 1.5$



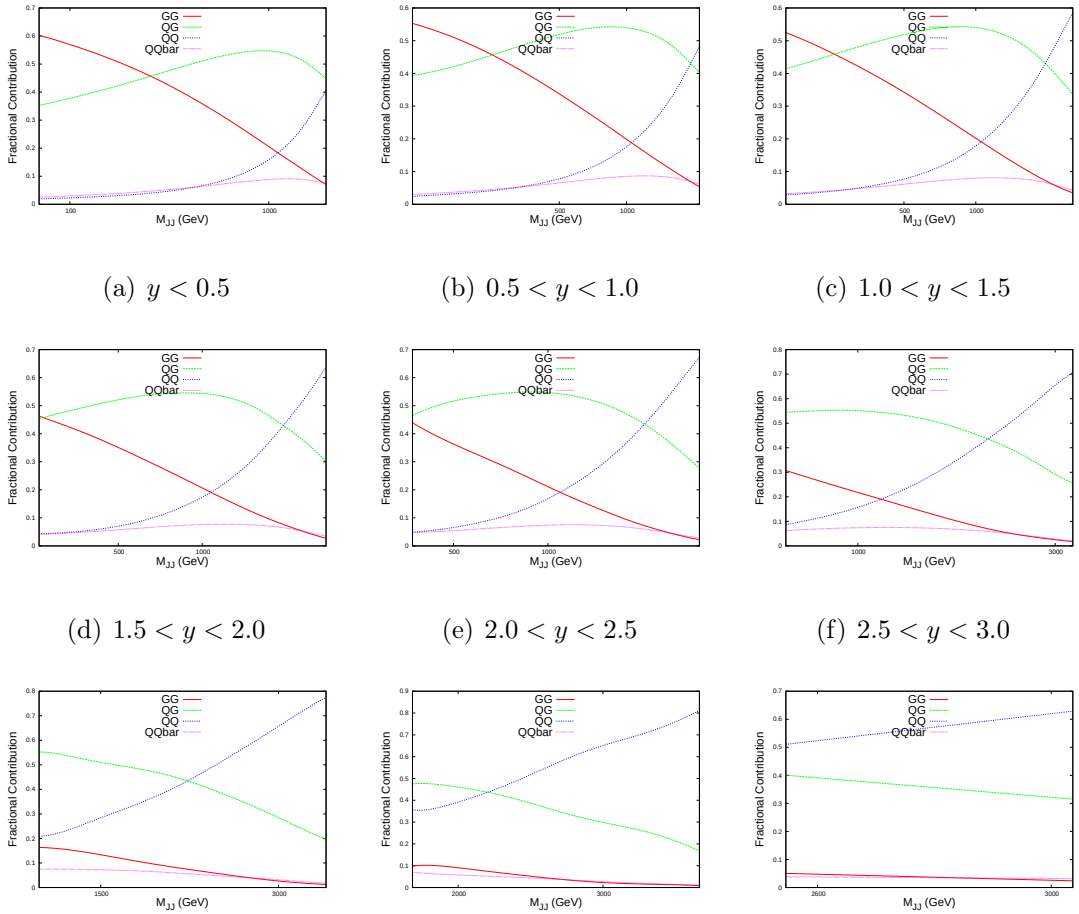
(d) $1.5 < y < 2.0$ (e) $2.0 < y < 2.5$ (f) $2.5 < y < 3.0$

Figure 4.4.: Contributions of different intial-state parton combinations to the D0 dijet cross section calculation

$$\frac{d^2\sigma}{dM_{JJ}dy} = \left[\alpha_s^2(\mu_R)\sigma_{LO} + \alpha_s^3(\mu_R)\left(\sigma_{NLO} + 2b_0 \log\left(\frac{\mu_R}{M_{JJ}}\right)\sigma_{LO} - 2\log\left(\frac{\mu_F}{M_{JJ}}\right)P_{ab}\otimes\sigma_{LO}\right)\right] \otimes f_a(\mu_F)\otimes f_b(\mu_F)$$

where the leading order and next to leading order cross sections, σ_{LO} and σ_{NLO} are computed using the matrix elements and evaluated at $\mu_R = \mu_F = M_{JJ}$, b_0 is the leading order QCD beta function coefficient, and P_{ab} are the QCD splitting functions. The behaviour of this cross section under renormalisation scale variations is relatively simple, with only the running of α_s and a logarithm including this variable. The factorisation scale variations, however, are sensitive to the convolution with the PDFs, and so the particular x values and partons probed in a particular event will affect the variations in μ_F.

Unlike inclusive jet production, in which the only physical scale involved in the events is the p_T of the jet, dijet production has a number of possible choices of scale. The seemingly most obvious choice is the average p_T of the two jets, however at high



(a) $y < 0.5$ (b) $0.5 < y < 1.0$ (c) $1.0 < y < 1.5$

(d) $1.5 < y < 2.0$ (e) $2.0 < y < 2.5$ (f) $2.5 < y < 3.0$

(g) $3.0 < y < 3.5$ (h) $3.5 < y < 4.0$ (i) $4.0 < y < 4.5$

Figure 4.5.: Contributions of different intial-state parton combinations to the ATLAS dijet cross section calculation

rapidities this can lead to problems due to the possible configuration of the event. A highly boosted hard scatter will have the same average p_T as an unboosted soft scatter. Another variable which could be used as the scale choice is the dijet mass, M_{JJ}, which would not suffer from the issues in event classification at high rapidities. At leading order, the mass is defined as:

$$M_{JJ} = 2p_T \cosh(y*) \tag{4.1}$$

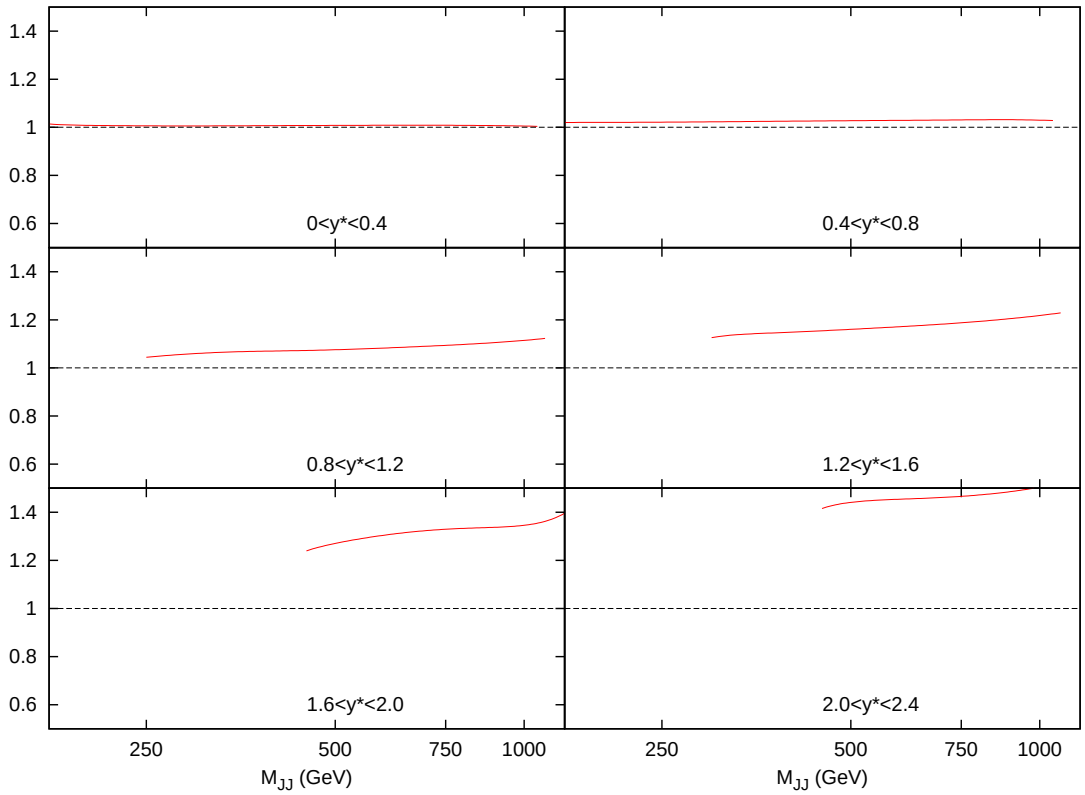


Figure 4.6.: Ratio of $\mu_R = \mu_F = M_{JJ}$ calculation to $\mu_R = \mu_F = 2 * p_T^{av}$ calculation for D0 dijet calculation. The equivalence of the two scale choices at central rapidities is apparent, with large deviations for more forward jets. Both calculations are performed with NLOjet++.

where $y* = (y_{jet1} - y_{jet2})/2$ is half the rapidity difference of the final state jets making the dijet pair. At the limit $y* = 0$, for fully back-to-back jets, we have $M_{JJ} = 2p_T^{av}$ as expected, and so the predictions using the two scale choices should agree. This is demonstrated in Fig 4.6, where the dijet cross section is calculated using both scales, and the ratio shown.

Fig 4.7 (in comparison to the p_T^{av} plot Fig 4.1) demonstrates the apparent benefit of using dijet mass as the scale choice. In the case of p_T^{av}, although at low rapidity the prediction is stable and flat across all M_{JJ}, the predictions from different multiplicative factors of the scale begin to cross in the more forward bins. This has already been observed in [68], however other scale choices were not investigated.

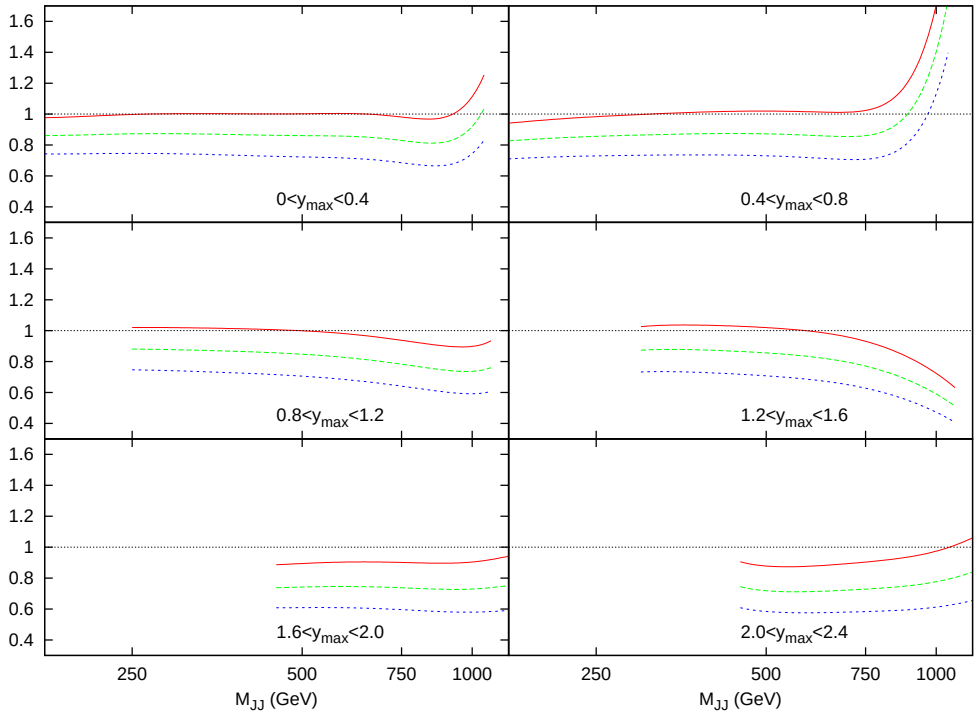


Figure 4.7.: Theory/Data ratio for D0 dijets, using multiples of M_{JJ} as the choice of μ_R and μ_F. The multiples are 0.5 (red), 1.0 (green) & 2.0 (blue)

In comparison, the theory/data ratio for the M_{JJ} calculation is much more stable. The variation through multiplicative factors of the scale are constant throughout all rapidity bins, and the ratio remains generally flat.

The χ^2 values shown in Tables 4.1−4.3 confirm that the choice of M_{JJ} provides the better fit to the D0 data. Also calculated is another choice of scale, namely multiples of $\frac{M_{JJ}}{2\cosh(0.7y*)}$. This form of scale choice was suggested by Soper et al. [77] as a means to stabilise NLO corrections, and is almost equivalent to the choice $p_T \exp(0.3y^*)$ used by ATLAS [52]. This choice allows the dependence on the dijet rapidity to be directly included. While it is also an improvement on the p_T^{av} calculation, it does not provide a better fit for the D0 dijets than using simply M_{JJ} alone.

The equivalent ATLAS results are now shown in Tables 4.4−4.6. The tendency for the p_T^{av} calculation to degrade at small multiplying factors is even more apparent

	$0.5 * p_T^{av}$	$1.0 * p_T^{av}$	$2.0 * p_T^{av}$
MSTW2008 NLO	3.23	2.34	1.61
CT10	4.79	4.08	2.74
NNPDF 2.1	2.12	2.20	1.21
ABKM	1.72	1.45	1.26
HERAPDF 1.5	2.01	2.00	1.79

Table 4.1.: χ^2 values for D0 dijets for different $\mu_R = \mu_F$.

	$0.5 * M_{JJ}$	$1.0 * M_{JJ}$	$2.0 * M_{JJ}$
MSTW2008 NLO	1.88	1.29	1.06
CT10	2.97	1.76	1.12
NNPDF 2.1	1.50	1.19	1.01
ABKM	1.14	1.25	1.38
HERAPDF 1.5	1.72	1.36	1.14

Table 4.2.: χ^2 values for D0 dijets for different $\mu_R = \mu_F$.

	$0.5 * \frac{M_{JJ}}{2\cosh(0.7y*)}$	$1.0 * \frac{M_{JJ}}{2\cosh(0.7y*)}$	$2.0 * \frac{M_{JJ}}{2\cosh(0.7y*)}$
MSTW2008 NLO	3.06	2.15	1.44
CT10	4.90	3.64	2.26
NNPDF 2.1	2.20	1.99	1.50
ABKM	1.66	1.38	1.12
HERAPDF 1.5	2.06	1.98	1.65

Table 4.3.: χ^2 values for D0 dijets for different $\mu_R = \mu_F$.

here than with the D0 dijets, so much so that the $0.5 * p_T^{av}$ is not shown, and all values are multiplied by a further factor of 2. Even with this additional factor, the $1 * p_T^{av}$ fit is poor for all PDF sets, and is due to the cross section calculation being negative in the high rapidity, high mass region, which can be seen in Fig 4.8. This plot clarifies the issue with using p_T^{av} that initially appeared in the D0 calculation, since it includes much higher rapidity and mass regions. It is clear that as higher rapidities are reached, the p_T^{av} calculation dramatically falls off for low multiplying factors, to the point where it becomes negative for both the 0.5 and 1.0 factors.

Despite this, once the multiplying factor is large enough, p_T^{av} provides the best fit of the three choices, with M_{JJ} in fact showing the worst fit of the three.

	p_T^{av}	$2.0 * p_T^{av}$	$4.0 * p_T^{av}$
MSTW2008 NLO	6.66	1.94	1.91
CT10	40.77	2.11	2.19
NNPDF 2.1	7.72	1.89	1.86
ABKM	10.35	2.01	2.17
HERAPDF 1.5	9.85	1.97	2.10

Table 4.4.: χ^2 values for ATLAS dijets for different $\mu_R = \mu_F$.

‘

	$0.5 * M_{JJ}$	$1.0 * M_{JJ}$	$2.0 * M_{JJ}$
MSTW2008 NLO	2.09	2.43	3.00
CT10	2.15	2.44	2.95
NNPDF	2.09	2.29	2.72
ABKM	2.43	3.11	3.91
HERAPDF	2.33	2.73	3.31

Table 4.5.: χ^2 values for ATLAS dijets for different $\mu_R = \mu_F$.

	$0.5 * \frac{M_{JJ}}{2\cosh(0.7y*)}$	$1.0 * \frac{M_{JJ}}{2\cosh(0.7y*)}$	$2.0 * \frac{M_{JJ}}{2\cosh(0.7y*)}$
MSTW2008 NLO	2.59	2.27	2.11
CT10	2.38	2.38	2.28
NNPDF 2.1	2.50	2.12	2.04
ABKM	2.19	2.28	2.44
HERAPDF	2.16	2.24	2.32

Table 4.6.: χ^2 values for ATLAS dijets for different $\mu_R = \mu_F$.

When considering the entire space of fits for any combination of (μ_R, μ_F), using the dijet mass is again shown to be a more stable prediction than p_T^{av}. Figs 4.9 and

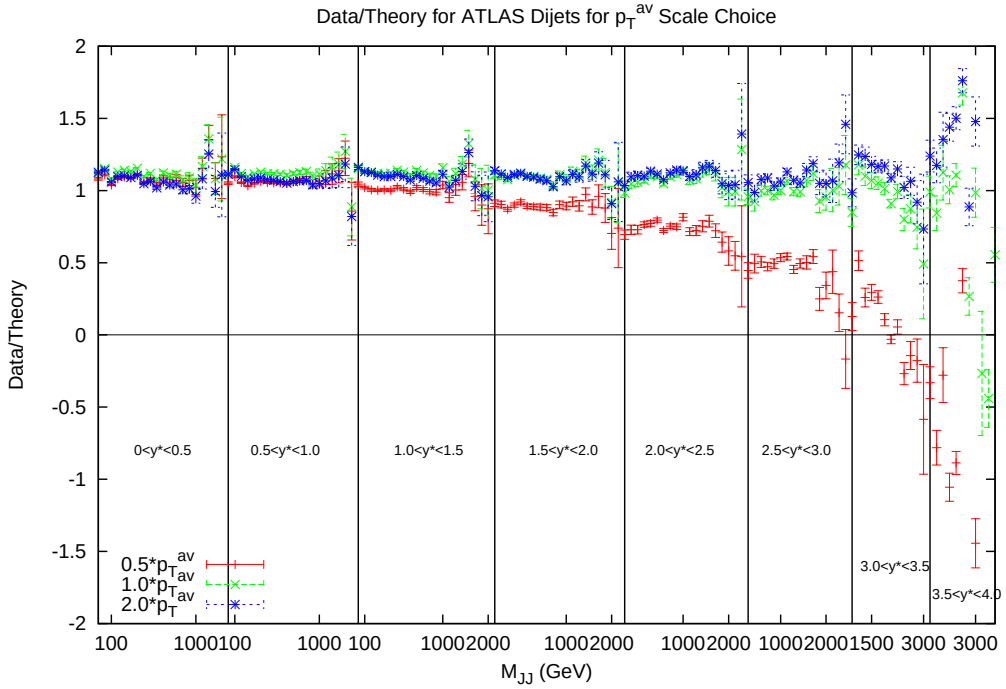


Figure 4.8.: Ratio of data to theory for ATLAS dijets using 3 different multiples of p_T^{av} as the scale choice. For the multiple of 1.0, the cross section becomes negative at high rapidity. This occurs much earlier for the lower multiple of 0.5.

4.10, which show the fit quality for D0 and ATLAS respectively, more completely shows the degradation of the p_T^{av} calculation at low values of scales. The yellow region, which for D0 covers the area in which either scale is below 0.5, shows a rapid unbounded increase in χ^2, deriving from the fact that the the cross section becomes increasingly negative as the scales approach 0. The fit becomes comparable in quality to the M_{JJ} calculation at much higher choices of scale, however there is no clear minimum which can be identified as a stable choice. For ATLAS, the region of divergent χ^2 is much larger for p_T^{av}, with normally sensible choices showing a very poor fit. Again, this is the result of the larger kinematic span of the ATLAS dijets exposing the failure of this calculation in the high rapidity, high mass region.

The M_{JJ} calculation for both data sets shows a similar trend by increasing towards lower scale choices. However, due to the stability at high rapidities, the fit

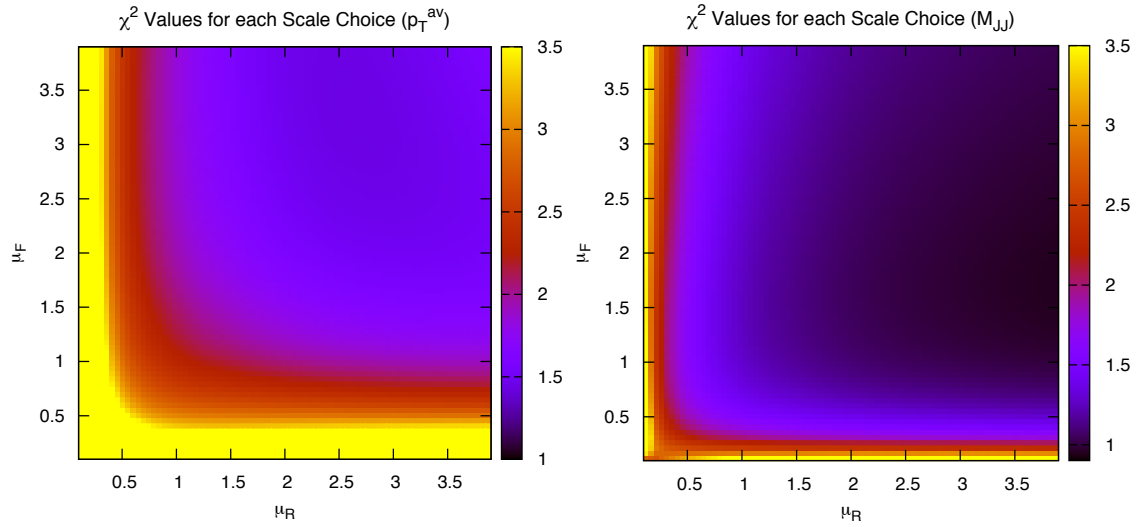


Figure 4.9.: χ^2 per point for all values of multiplication factor for both p_T^{av} and M_{JJ} calculations for D0 dijets. The yellow area at low scales in the p_T^{av} calculation is greatly off the scale, due to the calculation becoming negative in this region.

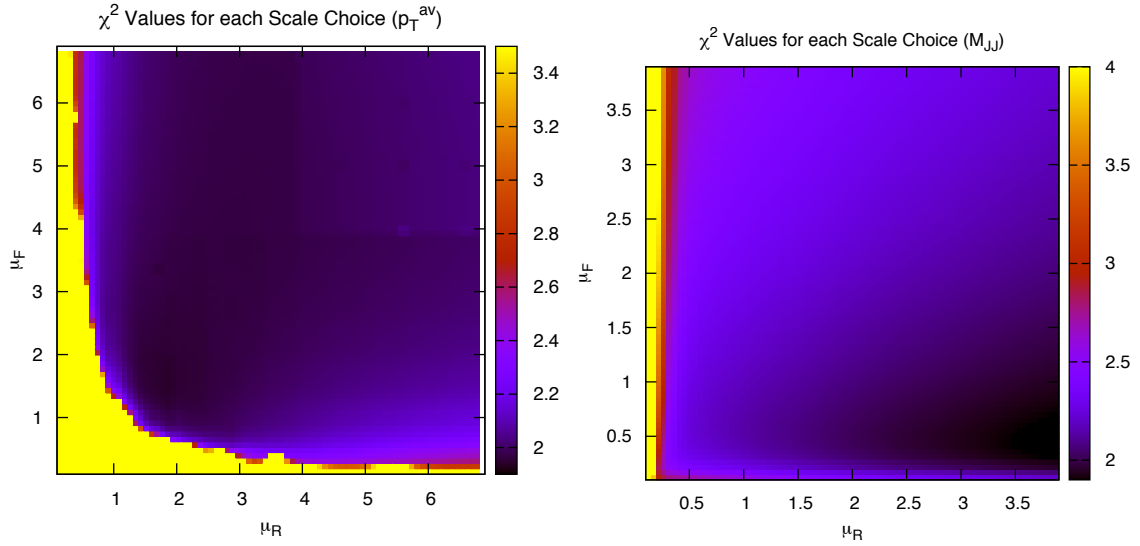


Figure 4.10.: χ^2 per point for all values of multiplication factor for both p_T^{av} and M_{JJ} calculations for ATLAS dijets. The yellow area at low scales in the p_T^{av} calculation is greatly off the scale, due to the calculation becoming negative in this region.

does not blow up in the same way as for p_T^{av}, and a lower χ^2 is apparent across the entire parameter space. There is a much clearer minimum, although it occurs for

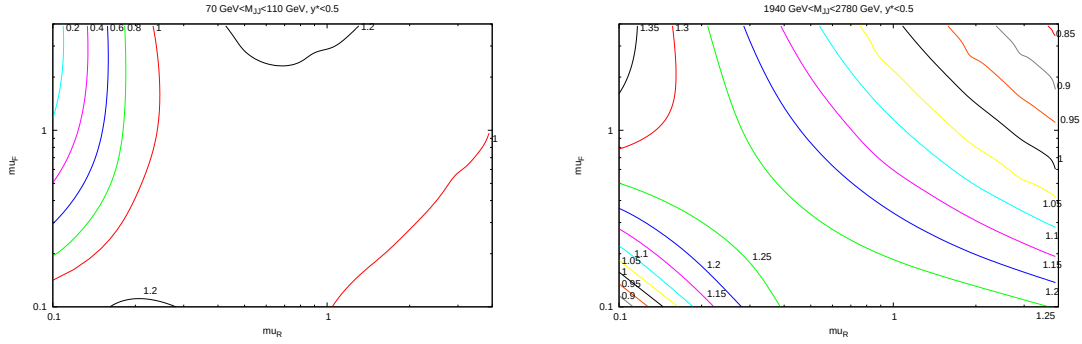


Figure 4.11.: Comparison of scale variations for the (a) lowest and (b) highest M_{JJ} bins in the $y* < 0.5$ rapidity bin of the ATLAS dijet calculation. The contour values are data/theory. The scales μ_R and μ_F are multiples of M_{JJ}

unusually high values of μ_R. This issue is discussed later and is shown to arise from the normalisation uncertainty.

4.2.1. Exclusive Cross Sections

The nature of the effect of scales can be more deeply probed by studying individual exclusive cross sections in finely defined regions of phase space. Whereas the previous discussion has focussed on the fit to data of an entire inclusive data set, the following will study the variation of each point within that data set for each scale choice. As Figs 4.4 & 4.5 have shown, the contributions from the inividual PDFs depends greatly on the values of the kinematic variables, and so the variation of each point in the factorisation scale direction should change in a similar manner.

Fig 4.11 demonstrates the scale variation of two single points in the kinematic phase space of the ATLAS data set. Both are in the lowest $y*$ bin, however the first includes dijets with low mass ($70 - 110$ GeV) and the second includes those with high mass ($1940 - 2780$ GeV). The general behaviour is that of a stable saddle region in the central region, with data/theory decreasing away from the saddle along one axis and increasing along the other. The axes defining the saddle region, however, differ greatly between the two points. A smooth rotation anticlockwise is observed as the dijet mass is increased, resulting in the large rotation shown in the figure.

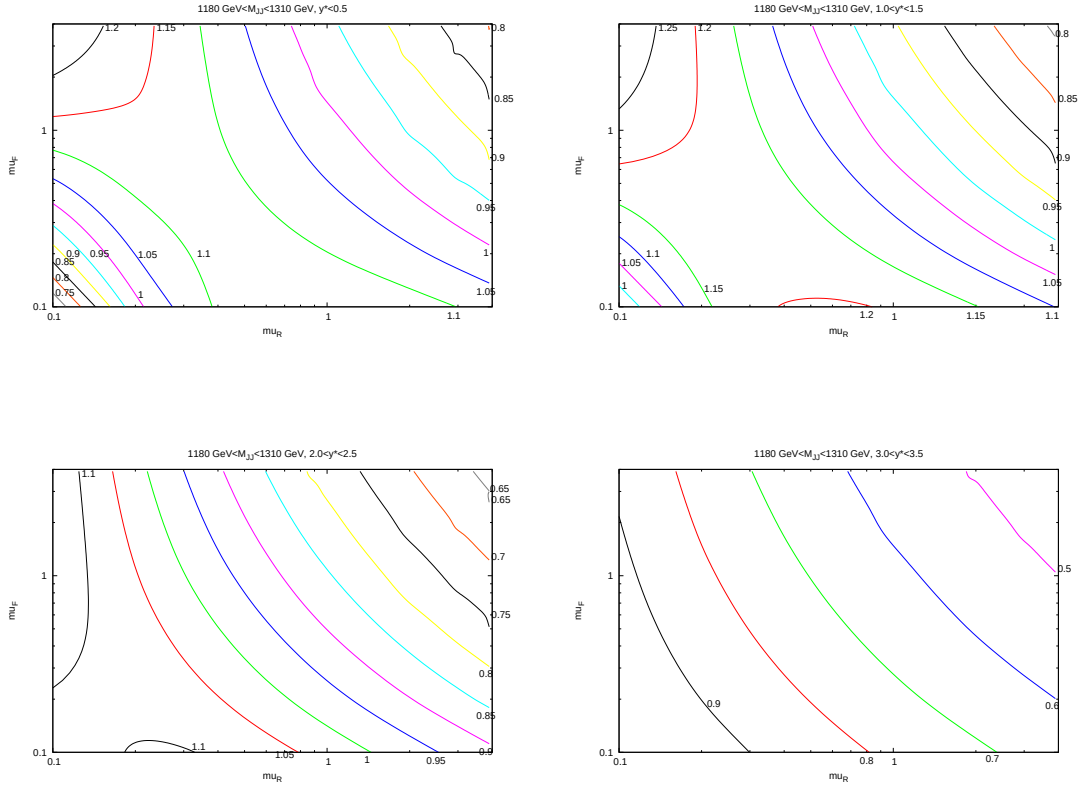


Figure 4.12.: Comparison of scale variations for different rapidity bins. The same M_{JJ} range is used throughout.

The dependence of this rotation on the kinematic variables is shown more clearly in Fig 4.12, where only the rapidity bin is changed. The 1.18 TeV $< M_{JJ} <$ 1.31 TeV bin is chosen for study as this is the bin appearing in the most rapidity bins. It is clear that the angle of the saddle point is dependent only on the dijet mass, however the overall behaviour is still affected by the rapidity. A migration towards lower scale choices is seen, such that at the highest rapidities, the saddle point disappears and the surface simply becomes a unidirectional slope. Ideally, the scale choice for a calculation would be thatwhich provides the most stable calculation, and hence would be within the saddle region for all of the points in the dataset. To demonstrate this, the contribution to the χ^2 from some selected individual points is shown for each normalisation value in Fig 4.13. For most points, the best fit corresponds to the saddle region, and there is a general inverse relationship between the fit and the slope away from the saddle.

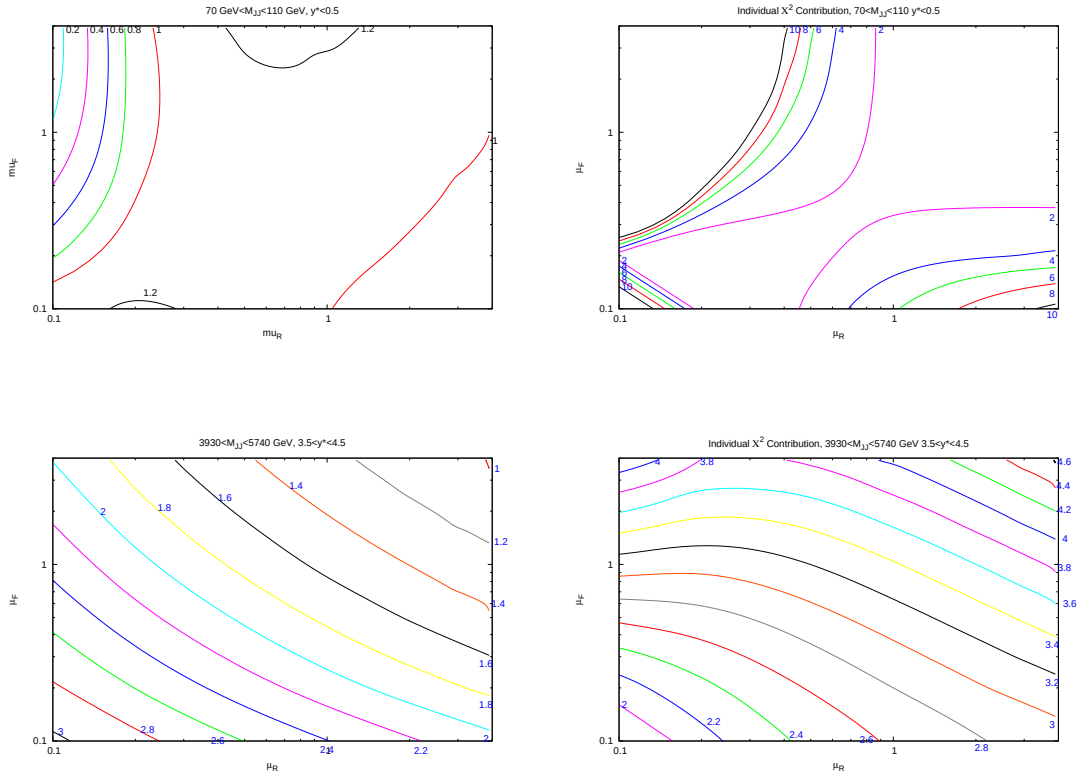


Figure 4.13.: The left hand plots show the data/theory contours for two different points, whilst the right hand plots show the contribution from the respective plots towards the total χ^2.

To understand the source of the observed behaviour, the variation in μ_R and μ_F are studied independently. Fig 4.14 demonstrates this for two points, at low and high M_{JJ} for the scale choice of M_{JJ}. The observed behaviour demonstrates that the rotation as a function of the dijet mass is governed by the factorisation scale changes. The renormalisation scale changes are similar at all values of M_{JJ}, with a smooth shape that changes little as the slices move through the factorisation scale range. The μ_F dependence, however, changes greatly with the dijet mass. In the first plot, with the lowest M_{JJ} bin in the lowest rapidity bin shown, the factorisation scale dependence is roughly flat for all slices in μ_R except for the very lowest 2 μ_R choices. This is the cause of the vertical nature of the saddle point in the first plot in Fig 4.11. In the second plot, at high M_{JJ} in the lowest rapidity bin, the factorisation scale has a non-flat shape that depends greatly on the value of μ_R chosen. Because the variations in factorisation scale are now large, the saddle point in the second

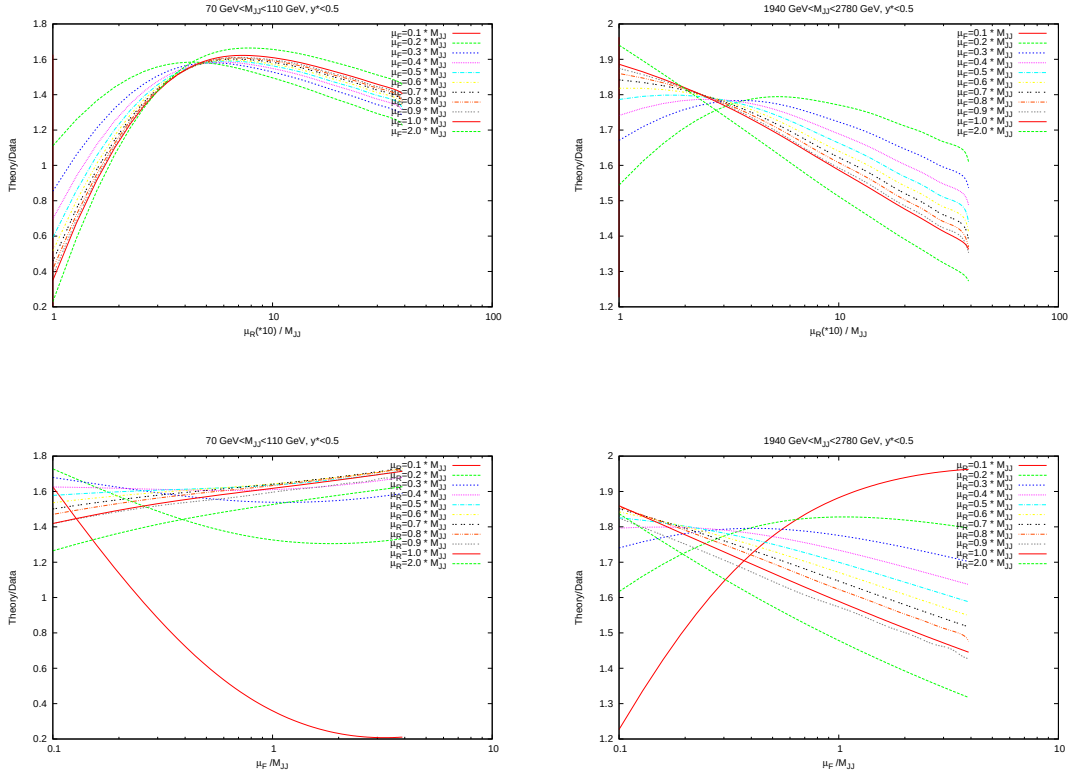


Figure 4.14.: Plots demonstrating the variation of μ_R and μ_F independently.

plot in Fig 4.11 is no longer vertical, and is rotated anticlockwise. The higher μ_F dependence at high M_{JJ} can be understood through the x values probed. In the high M_{JJ} region, the high x partons necessary for the events are evolved much more quickly than at low x, and so a greater dependence on the factorisation scale is seen. The stability of the calculation, then, is dependent on the partons probed.

The need to choose a single scale for the entire calculation leads to the search for a choice where the saddle point is uniformly based at that choice. Since the calculation using M_{JJ} as the kinematic scale choice seems to fail at higher rapidities, a function of M_{JJ} and $y*$ would be a logical choice to attempt to modulate this behaviour. The function $M_{JJ}/2\cosh(0.7y*)$ is studied, which was shown in the previous section to improve the stability of the ATLAS calculation. The scale variations for this choice are shown in Fig 4.15, where even in the highest rapidity bin, the saddle point is located around the central scale choice. It is clear that for the ATLAS dataset, the phase space probed requires a scale choice including a rapidity term.

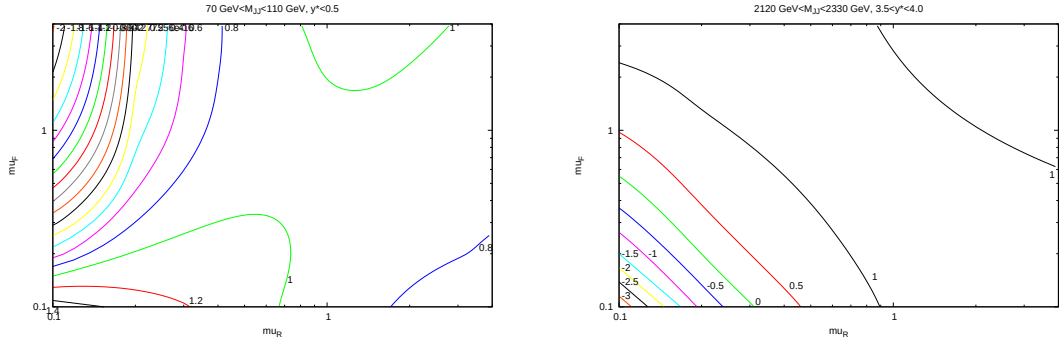


Figure 4.15.: Scale variations for the scale choice $M_{JJ}/2\cosh(0.7y*)$. Unlike when using M_{JJ}, the saddle point remains centrally located even in the high rapidity region.

By studying these plots of single data point scale variations, we can see that the stability of a calculation varies greatly when moving through phasespace, and so this should be considered in addition to the overall χ^2. The scale choice of $M_{JJ}/2\cosh(0.7y*)$ provides the best stability throughout the entire phasespace for dijet cross section calculations, whereas the other other two choices studied, p_T^{av} and M_{JJ}, do not demonstrate this favourable property.

4.3. Normalisation

The treatment of normalisation errors on data sets has been a subject of much discussion [80], and it is important to understand the effect they have on a fit. The only experimental source of the error is the luminosity uncertainty of the collider, and so it is correlated across all data sets produced at a single collider.

For the Tevatron Run II data, the luminosity uncertainty is 6.1%, whilst the ATLAS 7 TeV run has a 3.4% error. These provide the possibility for a theoretical prediction to move greatly up or down whilst incurring only a small penalty term in the χ^2. Due to this effect, the MSTW 2008 PDFs include a more severe quartic penalty term for the normalisation.

When considering the best choice of scale variable for D0 dijets, namely M_{JJ}, the best fit is obtained at very high values of renormalisation scale, as represented in

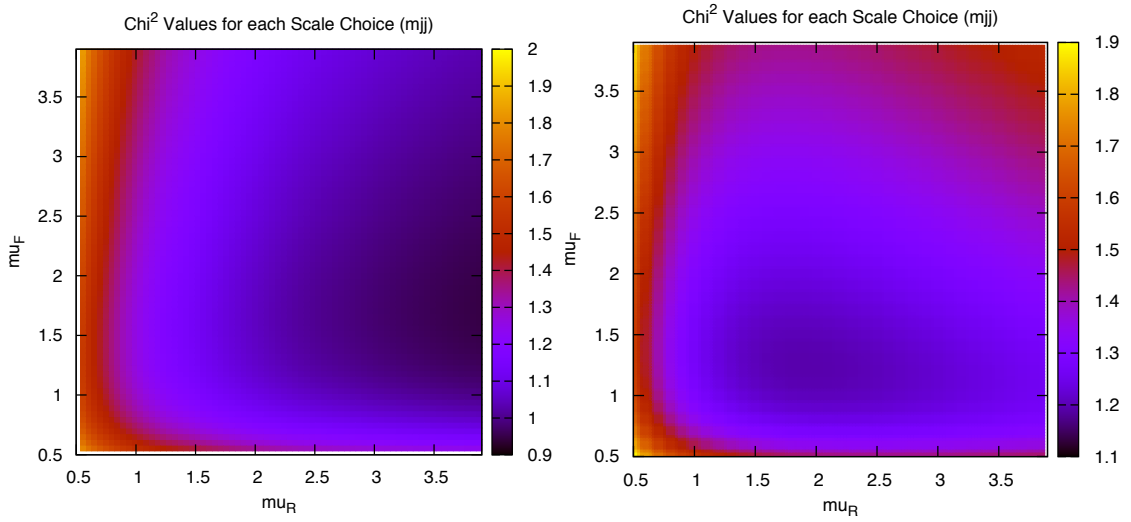


Figure 4.16.: Goodness of fit for each combination of scales (multiples of M_{JJ}), first with and second without allowing the normalisation to move freely.

Fig 4.16. However, if the normalisation r_k of each fit is studied, it is clear that this minimum is obtained in a region where a $2-3\sigma$ shift is required, as can be seen in Fig 4.17. In fact, there is a very small area of the parameter space in which the normalisation parameter is moved less than 1σ.

The second plot in Fig 4.16 represents the same fit, only keeping the normalisation fixed. The minimum is now at a more sensible scale choice, at the cost of requiring a slightly higher χ^2. Clearly equation 3.1 is inadequate for providing the most sensible fit, and a different treatment of the normalisation r_k is required. The difference in the normalisation treatments is most important for high values of the scales, where the calculation would naively appear to give the best fit.

4.4. Effect on MSTW PDFs

Fig 4.18 shows the change in the χ^2 for each eigenvector direction of the MSTW 2008 NLO set for the ATLAS dijet data, using 68% confidence levels. The plots show that, for the majority of the eigenvectors, a direction may chosen in which the fit quality may improve, if only slightly. The eigenvector which contributes most significantly across the inclusive- and dijet data sets is number 9, which is almost exclusively

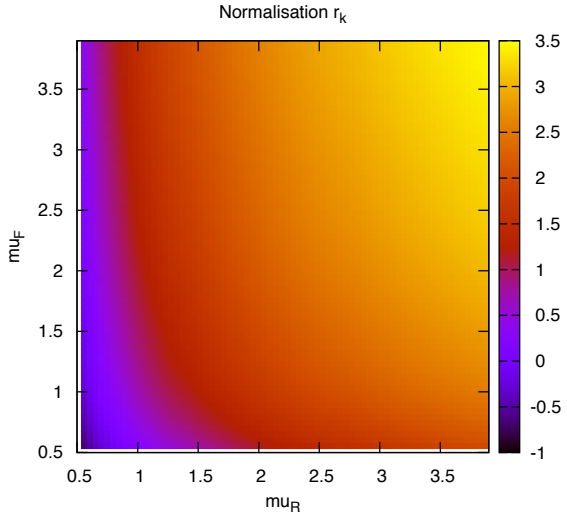


Figure 4.17.: Value of the systematic shift associated with the normalisation uncertainty for each scale value (multiples of M_{JJ}).

influenced by the gluon PDF. The other biggest contributors are influenced by a more mixed set of PDFs.

Next the reweighting procedure used in the previous chapter is repeated for the dijet datasets. The results for D0 dijets are shown in Fig 4.19. The scale choice used in the plots shown is M_{JJ}, however it was observed that a very similar effect was seen for the other two scale choices. Whilst the value of N_{eff} changes from 382 in the shown plots to 166 for p_T^{av} and 56 for $M_{JJ}/0.7\cosh(y*)$, the actual reweighted PDFs move in the same directions. All of the parton densities here are affected to some degree, with the largest deviations appearing around $x \sim 0.01$ where the Tevatron data has the most concentration. Notable is the fact that there is a reasonable shift from the central values, especially for the gluon which also sees an improvement in the error band at the previous noted x region. Given that the D0 inclusive jet data is included in the MSTW fit, this could be motivation to also attempt an inclusion of dijet data. The general trend of a larger gluon at low x and lower at high x, along with slightly larger quark densities overall is similar to that of the ATLAS and CMS inclusive jet data shown in the previous chapter.

Next, the PDFs are reweighted using the ATLAS dijet data. This time, a difference in the PDF effect is observed between the different scale choices, which indicates a fundamental difference in the implied physics. For the choice of M_{JJ}, shown in Fig 4.20, the gluon is moved well below its error band at moderate x values, and above

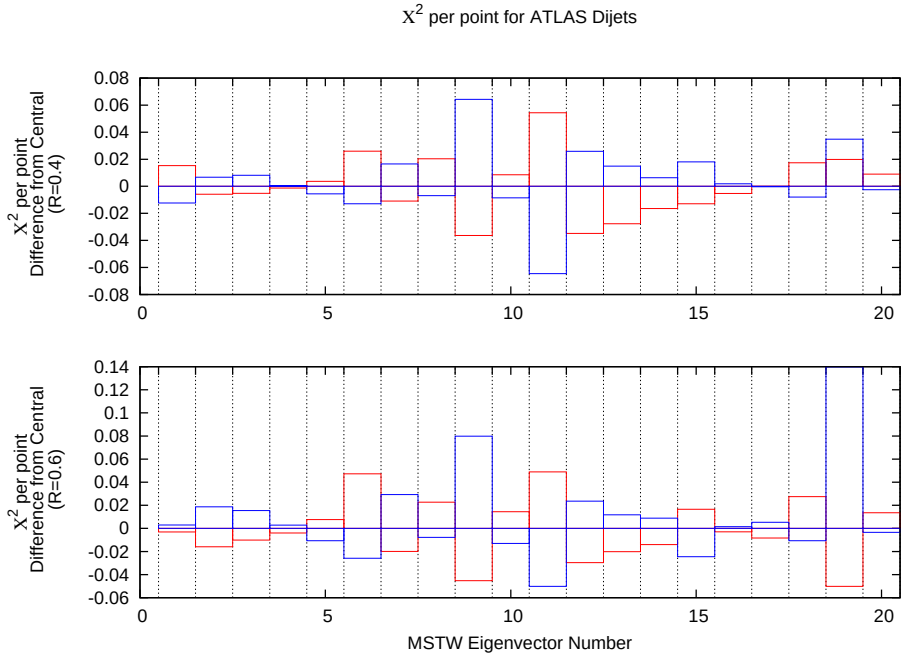


Figure 4.18.: Deviations in fit quality from the MSTW 2008 NLO central value for each of the 20 eigenvector directions. Blue (red) bars indicate the positive (negative) direction of deviations in the eigenvector dimension.

it at high x. All of the quark PDFs are also significantly shifted with a reduction in error band size. For the other two scale choices, shown in Figs 4.21 & 4.22, a less drastic and contradictory behaviour is seen, with the reweighted PDFs generally not moving outside of the error bands. These two choices also seem to agree well with each other, leaving M_{JJ} as the anomaly. All of the reweighted PDFs give an improved fit to data from the standard MSTW predictions: M_{JJ} changes from 2.30 to 1.95 per point, whilst p_T^{av} moves from 1.98 to 1.90. The value of N_{eff} is very low for the M_{JJ} and $M_{JJ}/0.7\cosh(y*)$ calculations, and so the results should be considered with due care. Any value below 100 implies that the reweighting is having a very large effect, and the original fit is very incompatible.

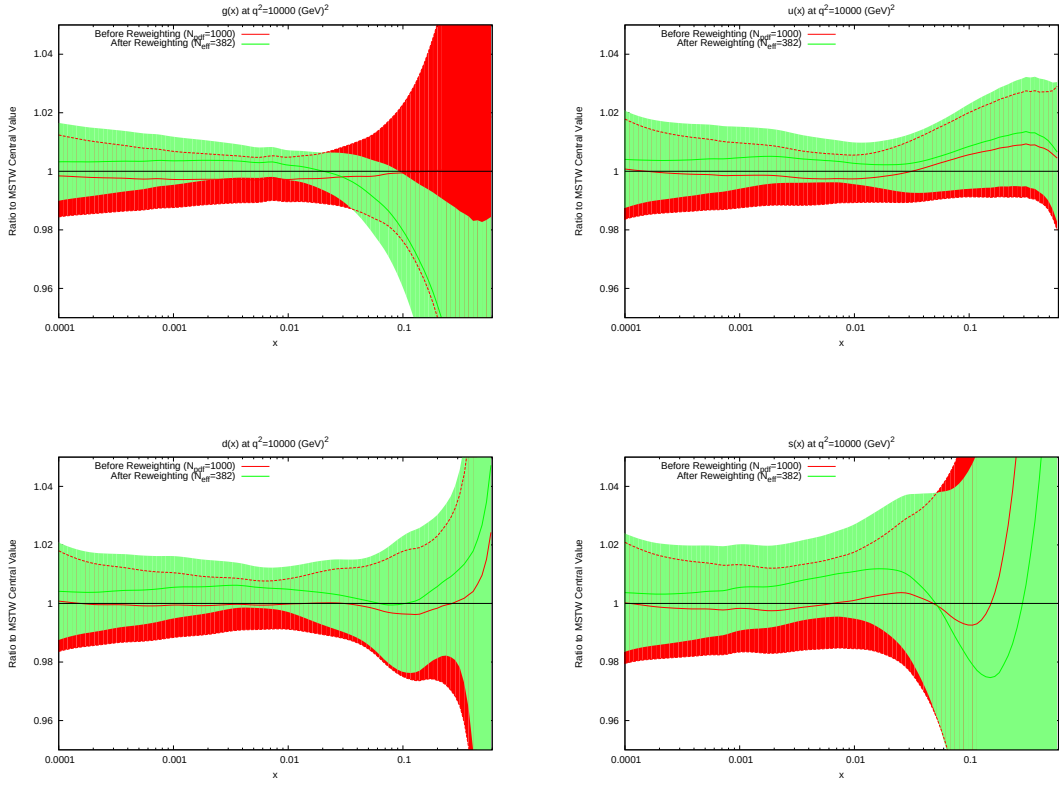


Figure 4.19.: Effect of PDF reweighting on the gluon, up, down and strange distributions for D0 dijet data.

4.5. CMS Dijets

Finally, as with the inclusive jets cross sections, the most recent and highest precision dijet data has come from the CMS experiment. The data consists of 54 points binned in M_{JJ} and y_{max}. This is significant since it is the same rapidity binning as D0, and different to ATLAS. Now any differences between the two approaches can be compared at the same collider. The x distributions of NLOjet++ events generated for this data set are shown for each rapidity bin in Fig 4.23. Due to the rapidity definition being the same as that at the Tevatron, the distribution resembles Fig 4.2, except with generally lower values of x probed. Here, central dijets are probed at around $x \sim 0.005$, with the highest rapidity dijets reaching $x \sim 0.0001$.

The data in this case extends less far in rapidity, from $y_{max} = 0$ to $y_{max} = 2.5$, than the ATLAS data, which went up to $y^* = 4.4$. Although the definitions are

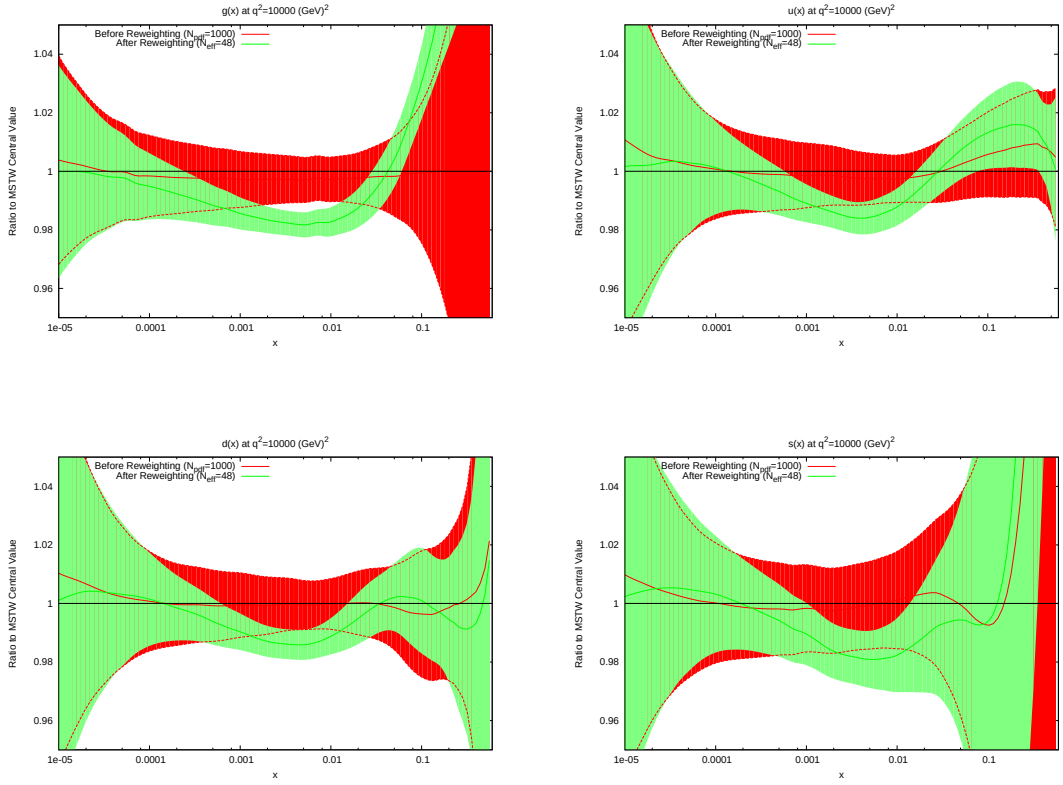


Figure 4.20.: Effect of PDF reweighting on the gluon, up, down and strange distributions for ATLAS dijet data. The scale choice used is M_{JJ}.

different, it must be true that the ATLAS data includes higher rapidity jets, since y^* is defined as half the difference of the dijets' rapidities, and so the highest bin necessarily only includes two very high rapidity jets.

The ratio of data to theory for the three scale choices is shown in Fig 4.24. The scale variation has less of an effect than for the ATLAS dijets, mostly due to the fact that the rapidity cut off is much lower, and the region where the most deviation occurred in the ATLAS dijets is avoided. The variation of the χ^2 fit with the scales for the p_T^{av} calculation is shown in Fig 4.25. Again, there is a region in the bottom left where the fit quality diverges exponentially, however this region is much smaller than the ATLAS dijets, again because of the lack of the high-rapidity region, where the calculation is known to go wrong. The results for all PDF sets are summarized in Table 4.7.

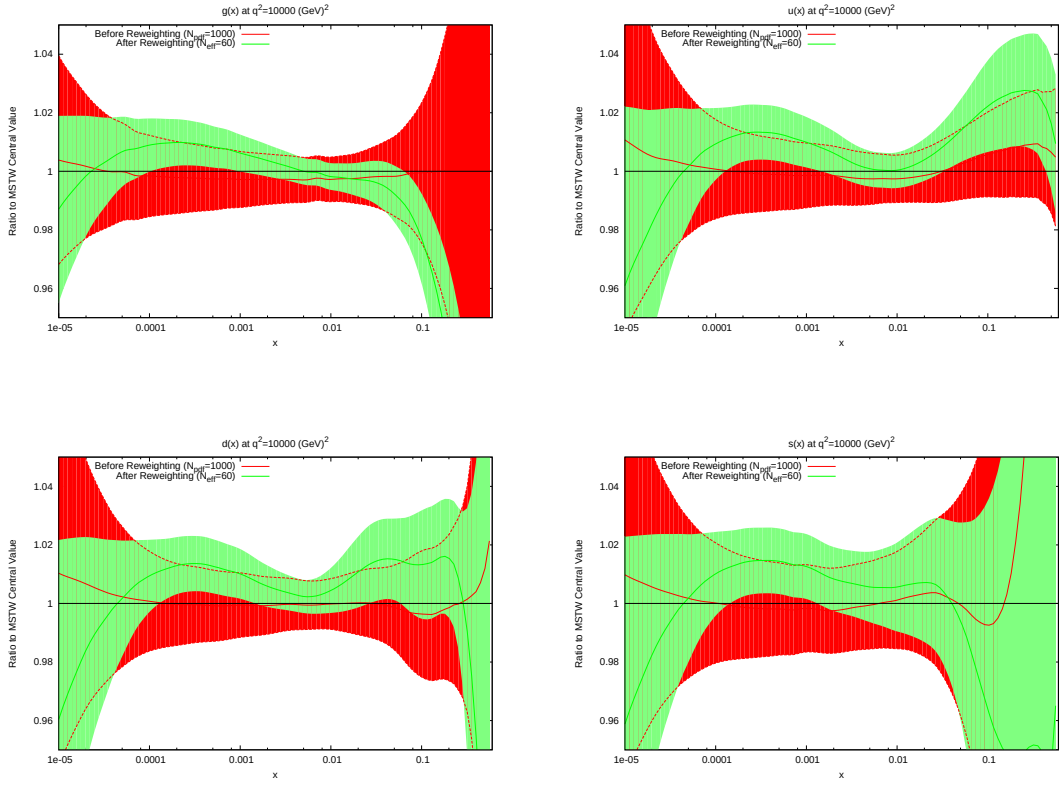


Figure 4.21.: Effect of PDF reweighting on the gluon, up, down and strange distributions for ATLAS dijet data. The scale choice used is $M_{JJ}/0.7\cosh(y*)$

	$0.5 * p_T^{av}$	$1.0 * p_T^{av}$	$2.0 * p_T^{av}$
MSTW2008 NLO	2.76	1.97	2.18
CT10	2.12	1.70	2.23
NNPDF	2.78	2.07	2.26
ABKM	3.80	3.31	4.56
HERAPDF	3.69	3.19	4.21

Table 4.7.: χ^2 values for CMS dijets

The results of the PDF reweighting are shown in Fig 4.26. Only the plot for the scale choice p_T^{av} are shown, since for this dataset the three choices are all in general agreement, unlike for the ATLAS dijets. The shape of the reweighted gluon is similar to that of the M_{JJ} ATLAS dijets, with a softer gluon at moderate x preferred. This effect is in contradiction to the preferred gluon of the inclusive jet data, implying a

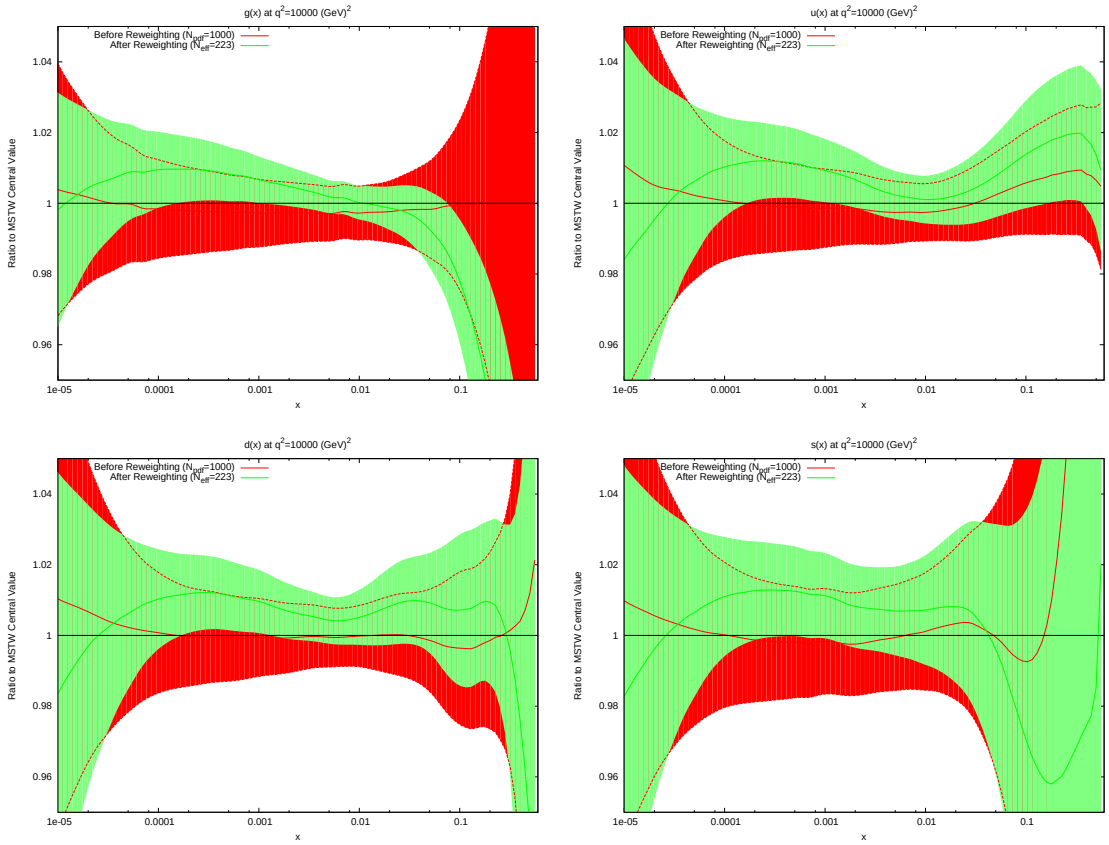


Figure 4.22.: Effect of PDF reweighting on the gluon, up, down and strange distributions for ATLAS dijet data. The scale choice used is p_T^{av}.

difference in the preferred PDF between the two datasets. This, in conjunction with the difference in preferred PDFs between scale choices, is further indication that the theoretical calculations in dijet production are not fully understood.

4.6. Summary

In this chapter, a detailed study into hadron-hadron dijet cross sections in relation to PDFs has been presented. The instability of the calculation observed at the Tevatron using the scale choice of p_T^{av} is explained by the behaviour of the kinematics at high rapidities. Calculations using other scale choices involving the dijet mass do not exhibit these problems, and so provide a more reliable estimate of the scale uncertainty.

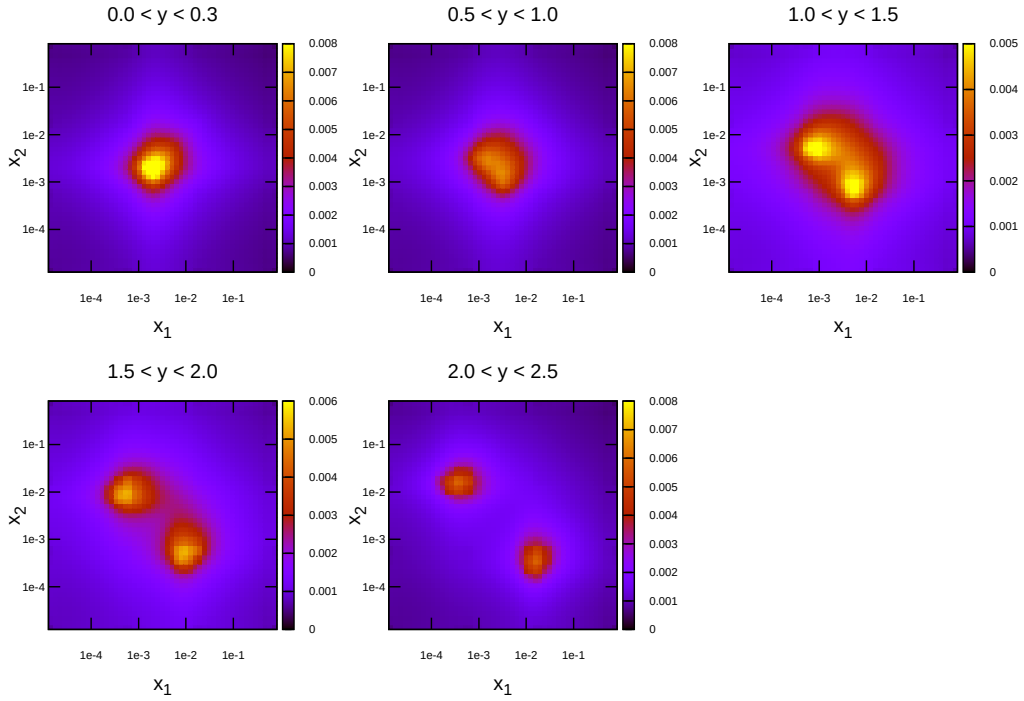


Figure 4.23.: Values of x_1 (highest x) and x_2 (lowest x) for each event generated in NLOJet++ for CMS dijets.

For ATLAS dijets, the instability is even more clear for the p_T^{av} calculation, with a very poor fit for low values of the scale multiplier quickly becoming an excellent fit for higher, unrealistic values. A study of the behaviour of the individual data points under scale variations demonstrates a saddle point structure which is centred around the central scale choice for low rapidity bins, and which can become a constantly decreasing plane at higher rapidities. The best scale choice to maintain the stability of each bin under scale variations is $M_{JJ}/0.7\cosh(y*)$.

The reweighting procedure has been conducted for each of three datasets, and in general the resulting preferred PDF depends upon the scale choice used. This is not an ideal situation, since the physics cannot depend on an unphysical mathematical property of the calculation. However, for the CMS dijet cross section, an agreement is reached between the scale choices, which is for a softer gluon across most of the x range, with the largest change at moderate x values. This also agrees with one

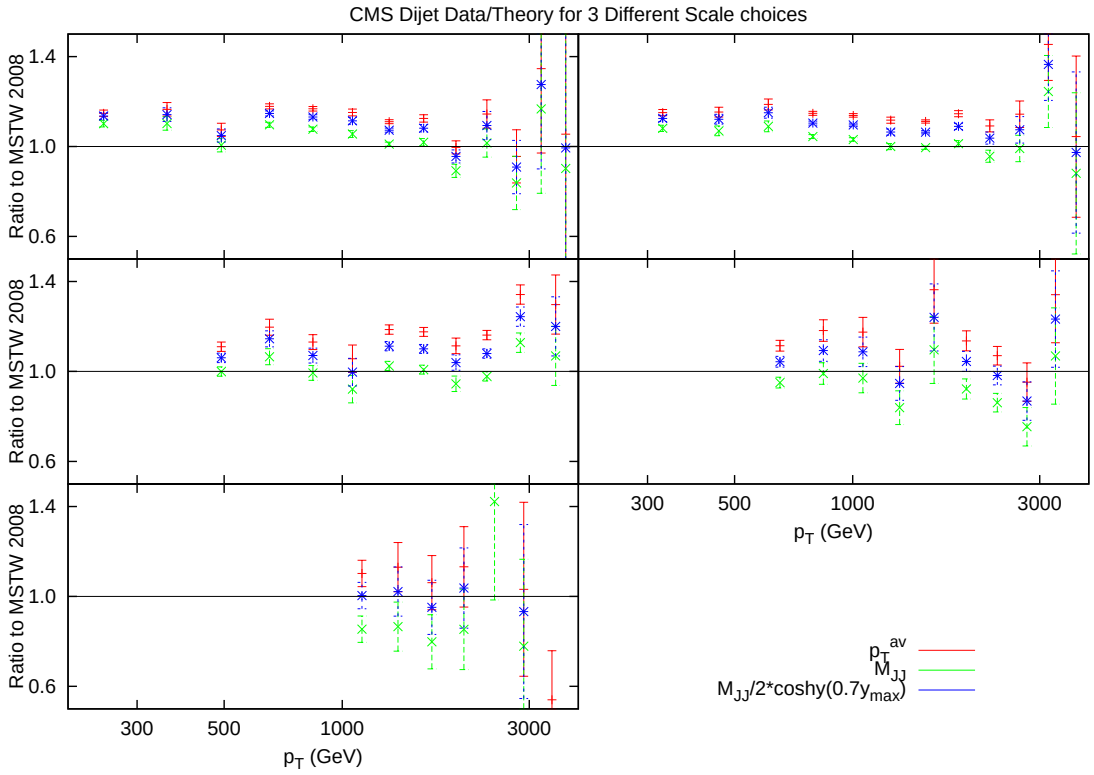


Figure 4.24.: Ratio of data to theory for CMS dijets for all rapidity intervals. All 3 of the scale choices discussed are shown.

of the scale choices for ATLAS dijets. This result is notable due to it being the opposite effect required to describe the ATLAS and CMS inclusive jet data, implying a conflict in the preferred PDFs for the two datasets.

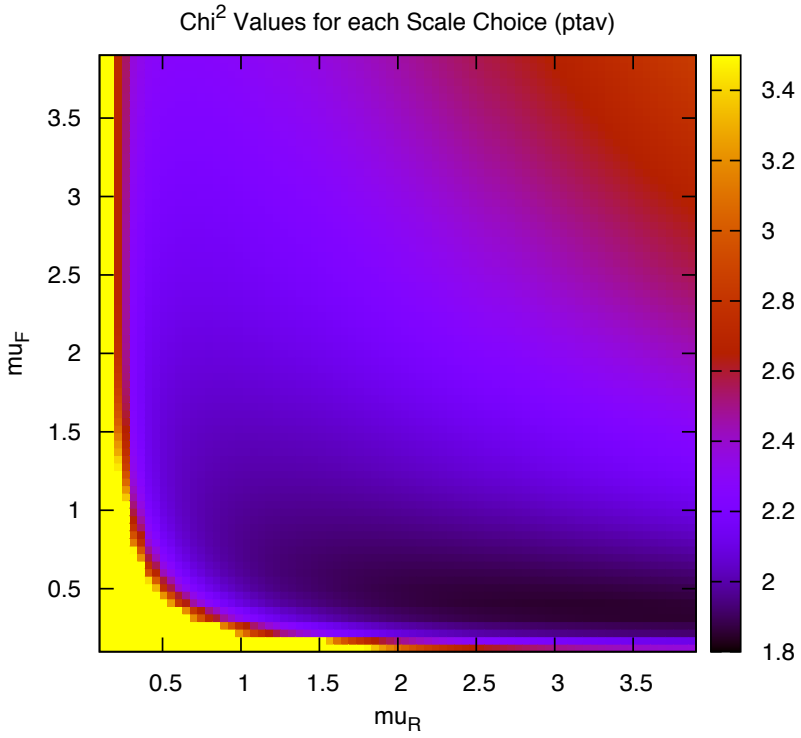


Figure 4.25.: χ^2 value for every combination of μ_R, μ_F for CMS dijets.

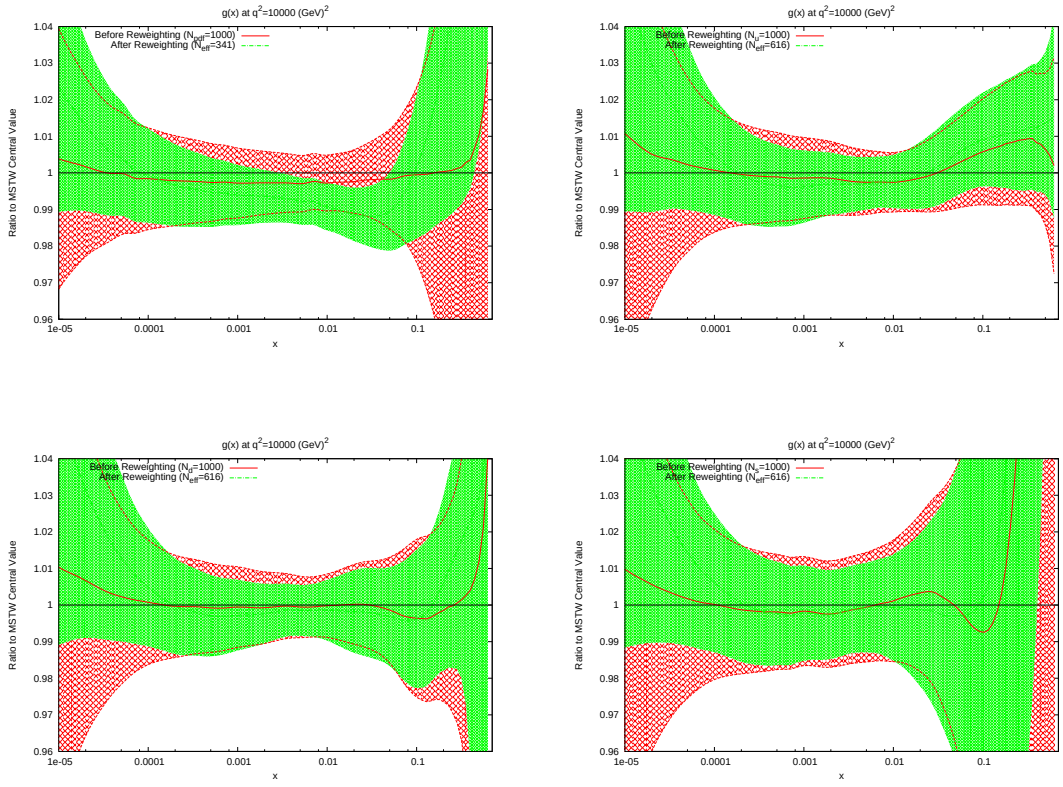


Figure 4.26.: Effect of PDF reweighting on the gluon, up, down and strange distributions for CMS dijet data. The scale choice used is p_T^{av}.

Chapter 5.

ATLAS W and Z Measurement

This chapter presents work conducted towards the paper [82] by the MSTW group. The PDF fits which are presented in the chapter were not work performed by the author, however the fit testing to ATLAS data sets and subsequent eigenvector reweighting are. A summary of the principles of the fit are presented first as necessary information.

The study of W and Z boson decays provides a unique window into the dynamics of certain PDFs, due to the initial state parton configurations probed. Specifically, when considering the W charge asymmetry, one can obtain direct information about the valence quarks. The asymmetry is defined as

$$A_W(y_W) = \frac{d\sigma(W^+)/dy_W - d\sigma(W^-)/dy_W}{d\sigma(W^+)/dy_W + d\sigma(W^-)/dy_W} \tag{5.1}$$

where y_W is the rapidity of the W boson. At leading order this can be shown to approximately equal a simple combination of quark PDFs:

$$A_W(y_W) = \frac{u_V(x_1)\bar{q}(x_2) + \bar{q}(x_1)u_V(x_2) - d_V(x_1)\bar{q}(x_2) - \bar{q}(x_1)d_V(x_2)}{u_V(x_1)\bar{q}(x_2) + \bar{q}(x_1)u_V(x_2) + d_V(x_1)\bar{q}(x_2) + \bar{q}(x_1)d_V(x_2) + 4\bar{q}(x_1)\bar{q}(x_2)} \tag{5.2}$$

where $x_{1,2} = (M_W/\sqrt{s})\exp(\pm y_W)$ are the momentum fractions.

The boson is not directly measured, however, and so the observable used is the rapidity of the charged lepton, which is similarly defined as:

$$A_W(y_l) = \frac{d\sigma(l^+)/dy_l - d\sigma(l^-)/dy_l}{d\sigma(l^+)/dy_l + d\sigma(l^-)/dy_l} \tag{5.3}$$

This complicates the picture of the underlying PDFs probed, since the angle θ^* of the lepton relative to the proton beam in the W boson rest frame must be additionally considered. This angle is dependent on the lepton's p_T, and is written as:

$$\cos^2\theta^* = 1 - 4p_T^2/M_W^2. \tag{5.4}$$

The rapidity of the lepton can then be related to the W boson rapidity through a function of this angle:

$$y_l = y_W + y^*,\ y^* = \frac{1}{2}\ln\left(\frac{1+\cos\theta^*}{1-\cos\theta^*}\right), \tag{5.5}$$

and now the leading order momentum fractions of the partons can be written as

$$x_{1,2} = x_0\exp(\pm y_W) = x_0\exp(\pm y_l)\kappa^{\pm 1},\ \kappa = \left(\frac{1+\cos\theta^*}{1-\cos\theta^*}\right)^{1/2} \tag{5.6}$$

With a knowledge of the parton momentum fractions and the form of the lepton coupling to the W boson, the numerator of the lepton asymmetry can be written as

$$\begin{aligned}&(u_v(x_1^+)\bar{q}(x_2^+) - \bar{q}(x_1^+)d_v(x_2^+) + u_v(x_1^-)\bar{q}(x_2^-) - \bar{q}(x_1^-)d_v(x_2^-))(1-\cos\theta^*)^2\\ &+(\bar{q}(x_1^+)u_v(x_2^+) - d_v(x_1^+)\bar{q}(x_2^+) + \bar{q}(x_1^-)u_v(x_2^-) - d_v(x_1^-)\bar{q}(x_2^-))(1+\cos\theta^*)^2\end{aligned} \tag{5.7}$$

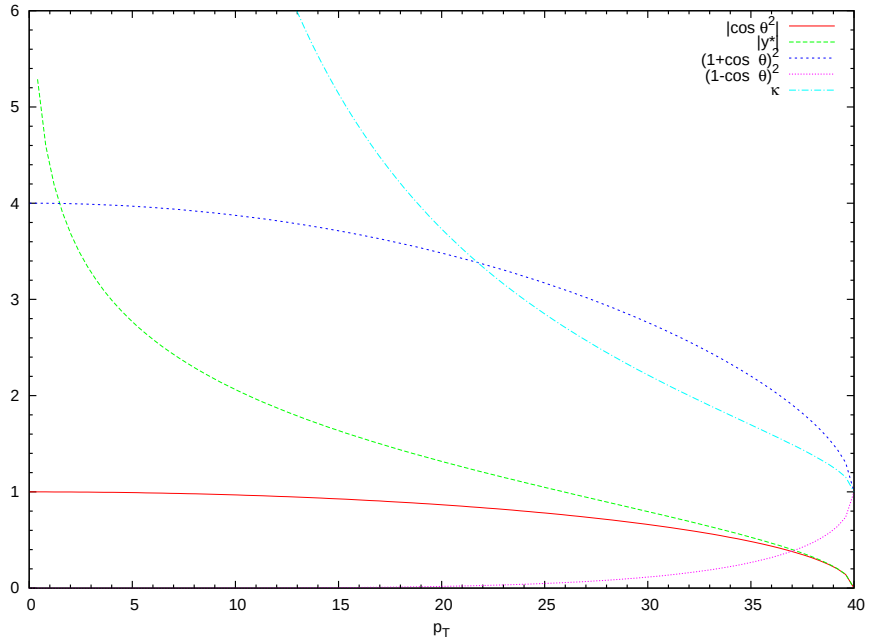


Figure 5.1.: The relevant lepton variables as a function of p_T.

The plot in Fig 5.1 shows the dependence on the various lepton variables on p_T. Most notable is the change in κ, which diverges for lower p_T values. Also of note is the difference between the $(1 + \cos\theta^*)^2$ and $(1 - \cos\theta^*)^2$ terms, which are similar at the W mass but become very different away from it. This demonstrates why the asymmetry itself is so dependent on p_T, as shown in Fig 5.2.

5.1. Reweighting Procedure

As with the jet cross sections in the previous chapter, the effect of the ATLAS W/Z measurements and the corresponding asymmetry data on MSTW PDFs is tested using the reweighting of PDFs randomly generated in eigenvector space. This study has been applied before in [73] for the CMS and ATLAS charged lepton asymmetry measurement, in which it was shown that the $u_v - d_v$ distribution required modification to properly describe the data. The result of reweighting in this case was to improve the χ^2 from 2.01 per point to 1.04 per point. The main cause of this improvement was shown to be the $u_v - d_v$ distribution, which was moved by the reweighting outside its error band.

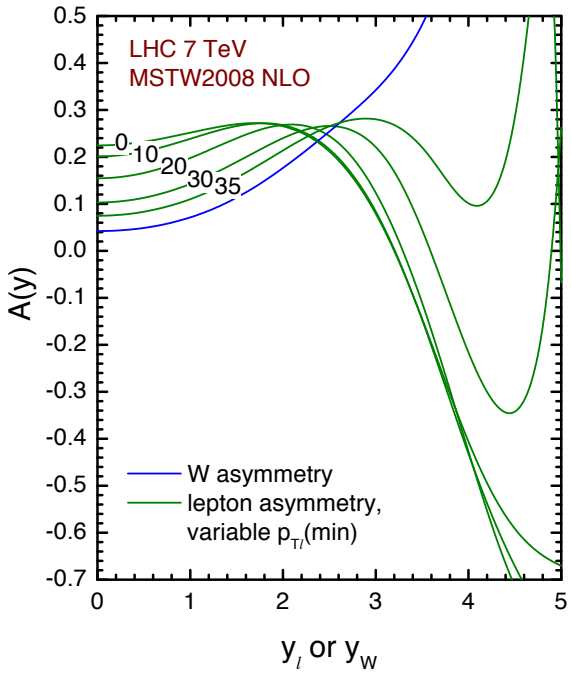


Figure 5.2.: The dependence of the lepton asymmetry on y and the lepton minimum p_T cut. The asymmetry is calculated at leading order and zero width using MSTW2008NLO PDFs. Plot taken from [82].

Here a study is conducted on the full W and Z cross sections published by the ATLAS collaboration [81]. The asymmetry data is a subset of this total cross section and loses information on the size of the total cross section and its shape in rapidity. The calculational tool used to produce the theoretical cross sections is again APPLgrid, this time using grids generated through MCFM. The systematic uncertainties are treated in the same manner as before, being incorporated into the χ^2 calculation Eq (3.1) and shifting the data points. The systematic uncertainties are correlated across all bins of the 3 separate cross sections, and so the shifts are well constrained. However, large shifts are unnecessary to achieve a good agreement between the raw data and theory, as shown in Fig 5.3, and for all PDFs only $\sim 25\%$ of the χ^2 originates from the penalty terms. The results for all major PDF sets are summarized in Table 5.1.

Next, the change in the prediction for each MSTW eigenvector is calculated, with the results shown in Fig 5.4. A large improvement is seen in eigenvectors 9,14 and 18, which are mainly affected by the gluon, u_v and d_v. The reweighting proedure

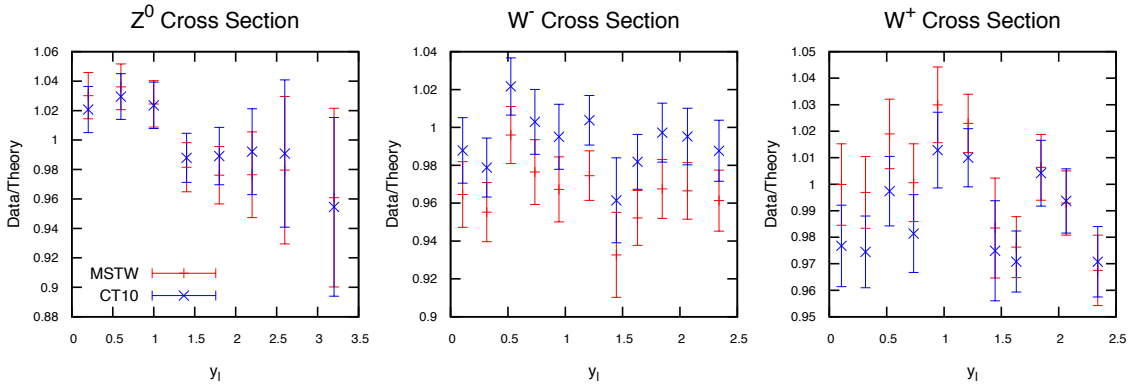


Figure 5.3.: Ratio of data to theory for the W and Z rapidity distributions. Only MSTW2008 and CT10 NLO PDFs are shown to avoid clutter. CT10 provides the best fit to the total cross section, with the difference most notable for the W^- cross section, where most points differ by the size of the error bands.

Scale	$M_{W/Z}/2$	$M_{W/Z}$	$2 * M_{W/Z}$
MSTW 2008	1.93	2.01	2.07
CT10	1.06	1.09	1.12
NNPDF 22	1.57	1.59	1.59
HERAPDF 15	1.63	1.68	1.73
ABKM09	1.57	1.59	1.59

Table 5.1.: Table of χ^2 per point (30 points) for NLO PDF sets for the W/Z total cross sections.

using these predictions is then applied, and the $u_v - d_v$ and gluon distributions are shown in Fig 5.5. There is little change in the gluon, however the $u_v - d_v$ requires a significant change in shape; smaller at low and high x with a hump at $x \sim 0.01$.

5.2. Chebyshev Polynomial PDFs

As discussed previously, one of the differences between the methods of PDF groups is the input parametrisation. In order to allow the low and high x PDFs to be properly affected by experimental data, a $(1-x)^n$ and an x^n (2 of these for the gluon) term are required. However, the form of the multiplicative polynomial which describes moderate x PDFs is not obvious to choose. One must allow enough looseness to

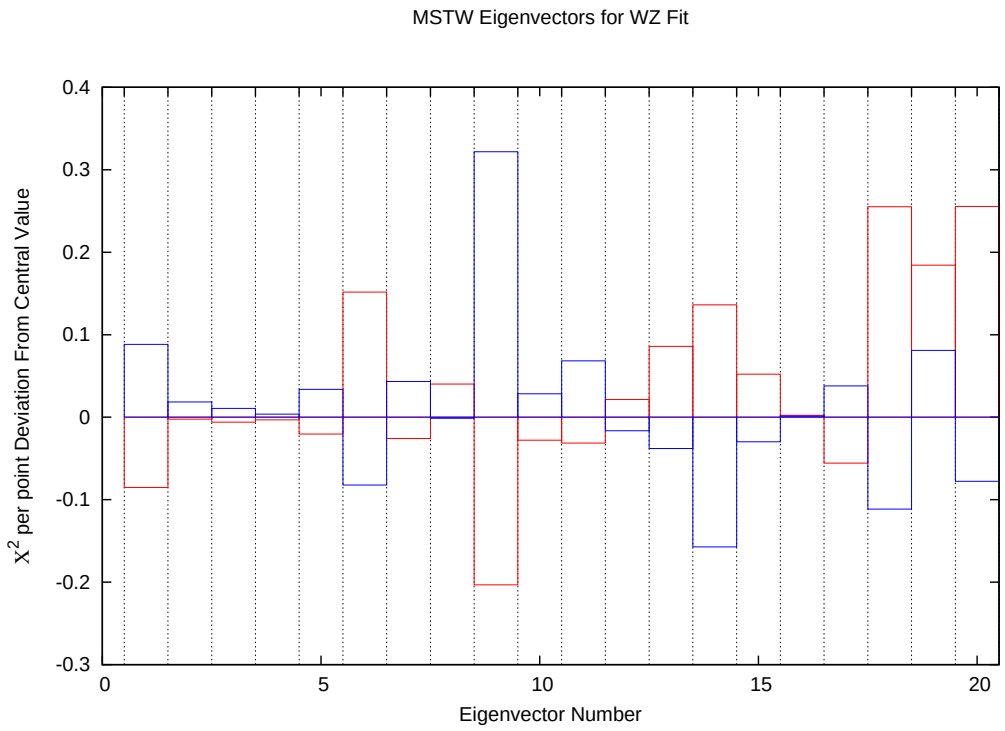


Figure 5.4.: Change in χ^2 fit for ATLAS W/Z cross section for each of the MSTW2008 eigenvectors.

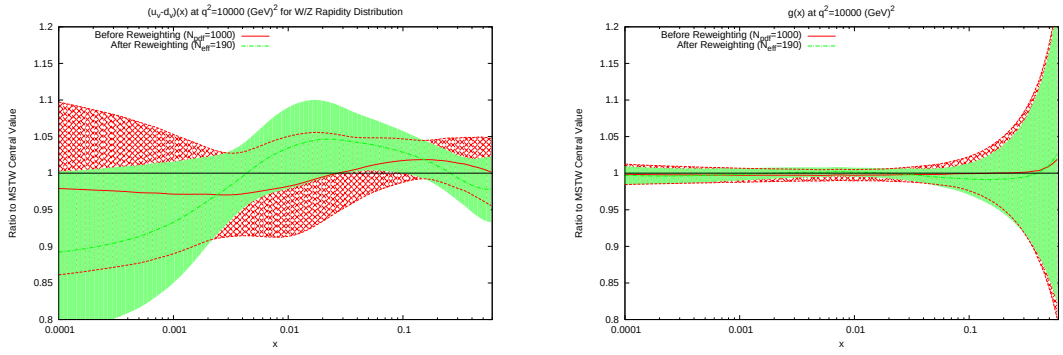


Figure 5.5.: The effect of parton reweighting on MSTW2008NLO PDFs using total W/Z ATLAS cross sections.

ensure the optimum fit is achieved, whilst also not allowing fluctuations in data to affect the fit. Quark distributions in the MSTW parametrisation are of the form:

$$xf_q(x, Q_0^2) = A(1-x)^{\eta_q} x^{\delta_q}(1 + \epsilon_q x^{0.5} + \gamma_q x^2), \tag{5.8}$$

where η_q, δ_q, ϵ_q and γ_q are the parameters allowed to vary in the fit process for the valence distribution.

The MSTW choice of half-integer powers in x is loosely motivated by Regge theory. Two coefficients were shown to be sufficient for the fit in previous studies, since inclusion of an extra term in the polynomial proportional to x^2 or $x^{0.25}$ were shown to not significantly improve the global fit to data, and most PDFs remain within their uncertainty bands. However, the u_v distribution is observed to move outside its error band. This fact, combined with the observation that u_v exhibited different behaviour using the Monte Carlo approach as opposed to varying the 28 MSTW parameters directly [73], led to the conclusion that the error on the u_v distribution could be underestimated.

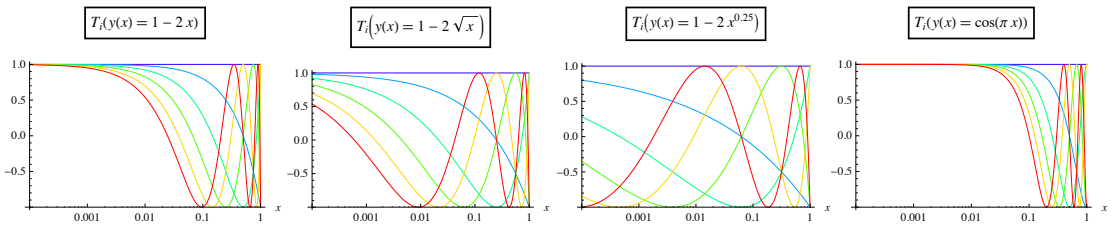


Figure 5.6.: Chebyshev polynomials of order i=0 to 5 for 4 different choices of expansion variable. The order increases from 0 for dark blue to 5 for red in each plot. Plot taken from [82].

Therefore, a complete study is undertaken into the form of the intermediate polynomial using Chebyshev polynomials, $T_i[y(x)]$, of various orders i. These are chosen because of some convenient properties, the most important of which is the fact that each maximum and minimum between the end points is equal to 1, and the end points themselves are equal to 1. This property allows the magnitude of variation to be the same across all values of x. The behaviour of the polynomials is shown in Fig 5.6, in which a selection of expansion functions $y(x)$ are shown for the first 6 orders. The new valence quark parametrisation is defined as

$$xf_q(x,Q_0^2) = A(1-x)^{\eta_q}x^{\delta_q}\left(1+\sum_{i=1}^{n}a_iT_i(y(x))\right), \tag{5.9}$$

where the high and low x terms are left the same as the standard MSTW parametrisation. For this analysis, the function is chosen to be $y(x) = 1-2\sqrt{(}x)$, which can be seen to give a good spread of possible values at all values of x. This is also consistent with the standard MSTW choice of half integer separation in the polynomials, which was motivated by Regge theory. The necessary order of the polynomial is shown in [82] to be 4.

The effect of these parametrisations on a global fit to data is shown in Figs 5.7 and 5.8. Initially only the valence quarks incorporate the Chebyshev polynomials (MSTWCPv), with only a minor improvement of 4 units in χ^2_{global}. The sea quarks are then additionally included in the extended parametrisation (MSTWCP), which produces a much larger improvement of 29 units. The main sources of improvement are BCDMS structure function data, whilst further slight improvements are seen for Tevatron lepton asymmetry and Z^0 data.

The largest effect on the parton distributions is seen in u_v, where both the CPv and CP PDFs are well outside the MSTW2008 error bands. There are changed in the other distributions, however these are generally within the error bands and become smaller with Q^2 evolution.

5.3. Deuteron Corrections & MSTWCPdeut Fit

For MSTW PDFs it is still necessary to include fixed target deuteron deep inelastic scattering to separate the valence quarks. This was previously the only data which had the ability to distinguish the u and d distributions. However with lepton asymmetry cross sections becoming available at both the Tevatron and LHC, it is possible to now probe these PDFs with collider data. In order to use the deuteron DIS data, one must include nuclear corrections to compensate for the deuteron target.

The standard MSTW choice in this regard is to include a linear negative correction at small x, but a subsequent study [83] which allowed a parametrised correction to

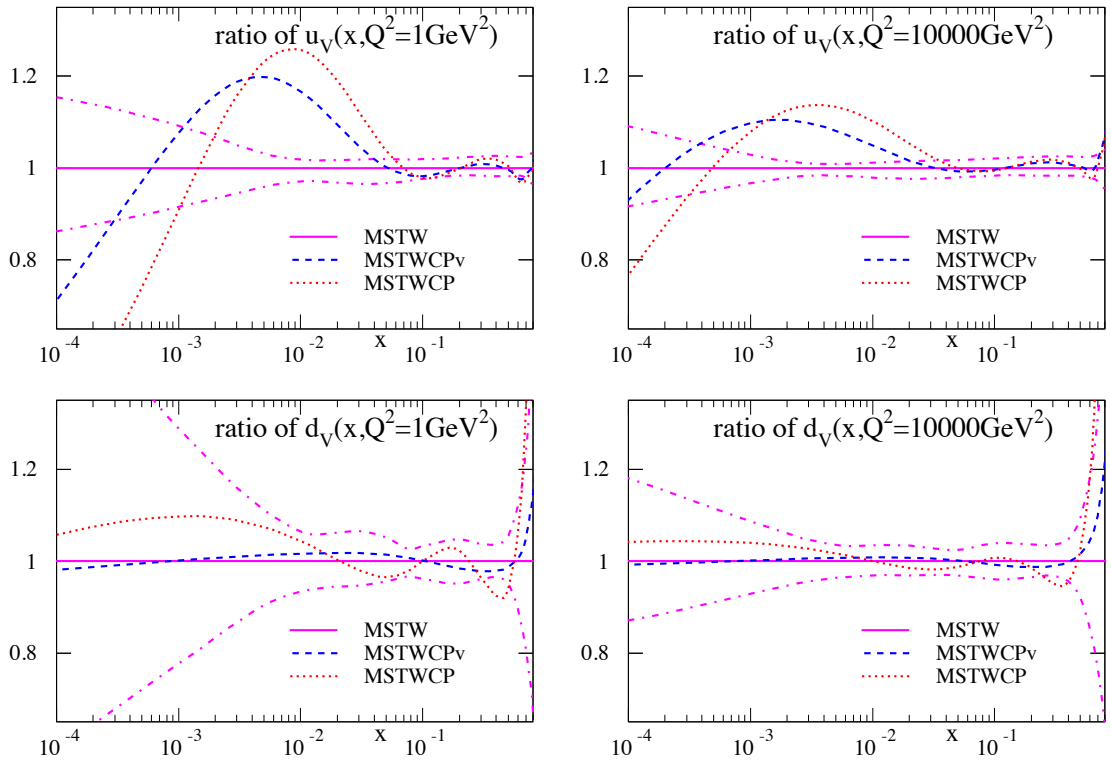


Figure 5.7.: Effect of Chebyshev parametrisation on MSTW valence quark PDFs. Only valence quarks are included in the extended parametrisation for MSTWCPv, and sea quarks are added for MSTWCP. Plot taken from [82].

select the optimum choice, found a form of a negative dip at around $x \sim 0.3$, followed by a verge large positive correction above this. Since the Chebyshev polynomial parametrisation shows a large valence quark correction, it is natural to test the new deuteron corrections on these PDFs.

Whilst the previous study found an improvement in global χ^2 of approximately 80 units from the MSTW2008 fit, the Chebyshev polynomial PDFs have a slightly larger reduction of 86 units. The contributions to the χ^2 from each of the data sets is shown in Table 5.2. The most noticeable improvement in the MSTWCP PDFs from the MSTW2008 set are in the BCDMS μd data set, which is most affected by the valence quark distributions. The largest improvement seen when moving from MSTWCP to MSTWCPdeut is in the Tevatron lepton asymmetry data. This can be understood through the improved $u_v - d_v$ constraint in these PDFs, which is the most important distribution for lepton asymmetry.

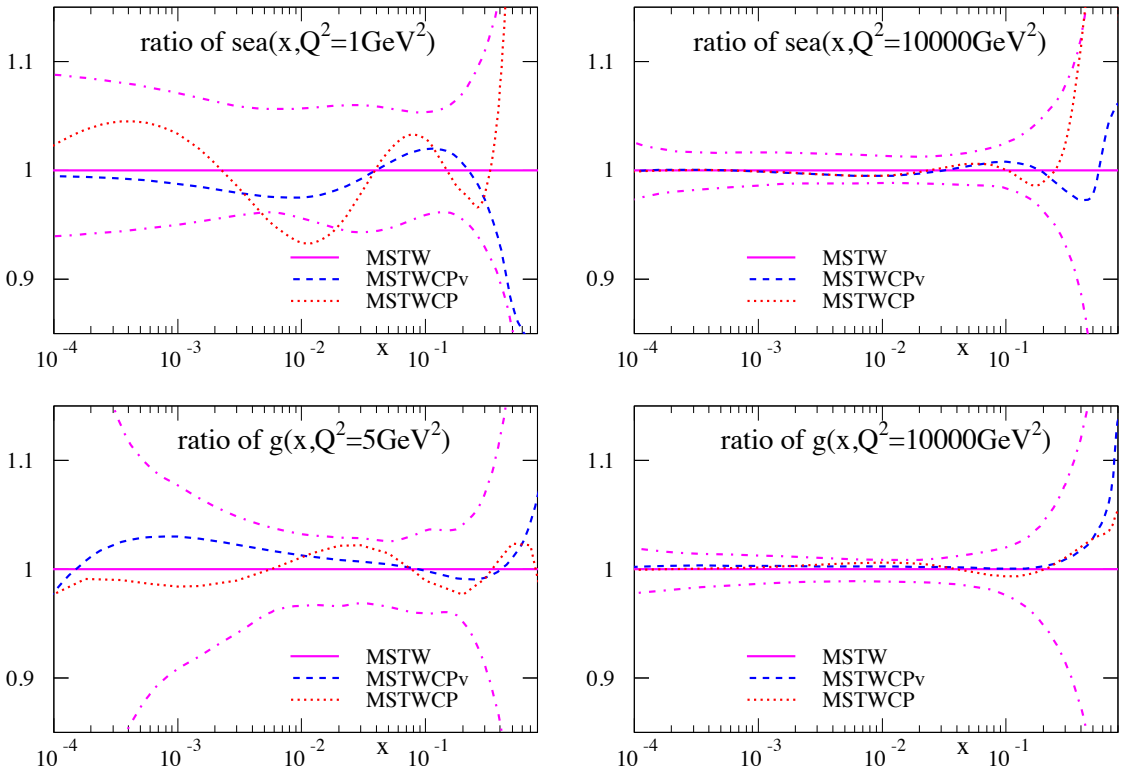


Figure 5.8.: Effect of Chebyshev parametrisation on MSTW sea quark and gluon PDFs. Only valence quarks are included in the extended parametrisation for MSTWCPv, and sea quarks are added for MSTWCP. Plot taken from [82].

Fig 5.10 shows the change in the valence quark distributions for the MSTWCPdeut and MSTWCP fits compared to MSTW2008. For the up valence quark, both of the new fits provide a greatly different shape than the 2008 fit, with the majority of the x range showing shifts outside of the error bands. MSTWCP is generally in agreement with MSTW2008 for the down valence quark, however the additional deuteron corrections in MSTWCPdeut require a higher low-x and lower moderate-x distribution.

Data set	MSTW08	MSTWCP	MSTWCPdeut
BCDMS μp F_2	182 / 163	173 / 163	177 / 163
BCDMS μd F_2	190 / 151	168 / 151	143 / 151
NMC μp F_2	121 / 123	123 / 123	120 / 123
NMC μd F_2	102 / 123	101 / 123	103 / 123
NMC $\mu n/\mu p$	130 / 148	143 / 148	143 / 148
E665 μp F_2	57 / 53	55 / 53	53 / 53
E665 μd F_2	53 / 53	58 / 53	57 / 53
SLAC ep F_2	30 / 37	31 / 37	31 / 37
SLAC ed F_2	30 / 38	31 / 38	31 / 38
NMC/BCDMS/SLAC F_L	38 / 31	39 / 31	39 / 31
E866/NuSea pp DY	228 / 184	224 / 184	221 / 184
E866/NuSea pd/pp DY	14 / 15	10 / 15	7 / 15
NuTeV νN F_2	49 / 53	49 / 53	54 / 53
CHORUS νN F_2	26 / 42	25 / 42	26 / 42
NuTeV νN xF_3	40 / 45	48 / 45	45 / 45
CHORUS νN xF_3	31 / 33	34 / 33	32 / 33
CCFR $\nu N \to \mu\mu X$	66 / 86	65 / 86	64 / 86
NuTeV $\nu N \to \mu\mu X$	39 / 40	38 / 40	39 / 40
H1 MB 99 e^+p NC	9 / 8	8 / 8	8 / 8
H1 MB 97 e^+p NC	42 / 64	40 / 64	40 / 64
H1 low Q^2 96–97 e^+p NC	44 / 80	44 / 80	44 / 80
H1 high Q^2 98–99 e^-p NC	122 / 126	121 / 126	120 / 126
H1 high Q^2 99–00 e^+p NC	131 / 147	129 / 147	129 / 147
ZEUS SVX 95 e^+p NC	35 / 30	35 / 30	35 / 30
ZEUS 96–97 e^+p NC	86 / 144	87 / 144	87 / 144
ZEUS 98–99 e^-p NC	54 / 92	53 / 92	53 / 92
ZEUS 99–00 e^+p NC	63 / 90	62 / 90	61 / 90
H1 99–00 e^+p CC	29 / 28	28 / 28	31 / 28
ZEUS 99–00 e^+p CC	38 / 30	38 / 30	35 / 30
H1/ZEUS ep F_2^{charm}	107 / 83	108 / 83	108 / 83
H1 99–00 e^+p incl. jets	19 / 24	19 / 24	19 / 24
ZEUS 96–97 e^+p incl. jets	30 / 30	29 / 30	29 / 30
ZEUS 98–00 $e^\pm p$ incl. jets	17 / 30	16 / 30	16 / 30
DØ II $p\bar{p}$ incl. jets	114 / 110	117 / 110	113 / 110
CDF II $p\bar{p}$ incl. jets	56 / 76	57 / 76	56 / 76
CDF II $W \to l\nu$ asym.	29 / 22	26 / 22	18 / 22
DØ II $W \to l\nu$ asym.	25 / 10	20 / 10	9 / 10
DØ II Z rap.	19 / 28	18 / 28	17 / 28
CDF II Z rap.	49 / 29	45 / 29	52 / 29
All data sets	**2543 / 2699**	**2513 / 2699**	**2457 / 2699**

Table 5.2.: Table of χ^2 values for the two new fits, MSTWCP and MSTWCPdeut, compared to the standard MSTW2008NLO fit. Plot taken from [82].

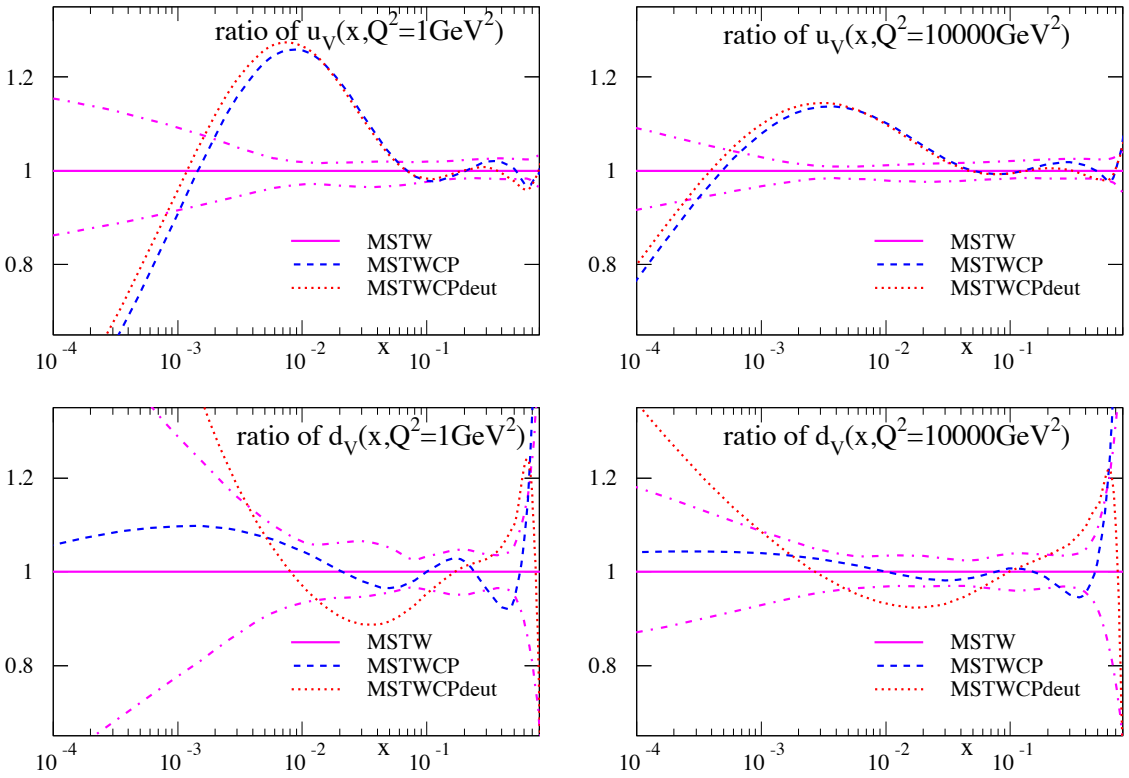
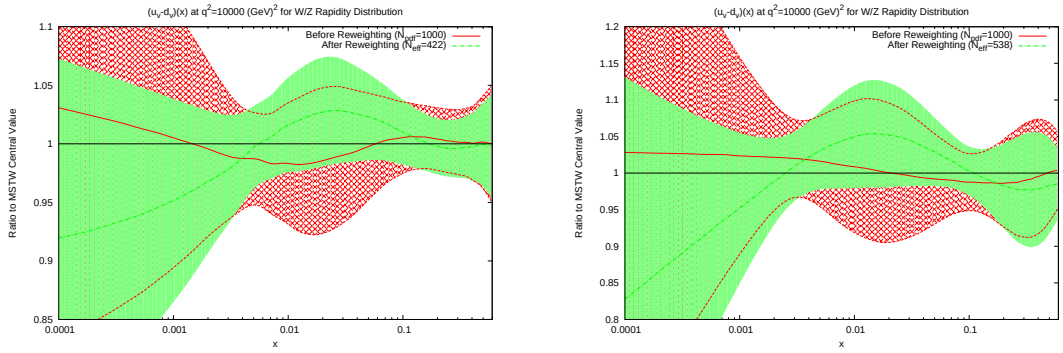


Figure 5.9.: Ratio of the Chebyshev polynomial plus deuteron corrections to standard MSTW2008 NLO PDFs with 68% uncertainty. Plot taken from [82].



(a) MSTWCP (b) MSTWCPdeut

Figure 5.10.: Effect of PDF reweighting on the $u_v - d_v$ distribution for MSTWCP and MSTWCPdeut PDFs.

5.4. W/Z Predictions Using MSWCP & MSTWCPdeut

Now we return to the ATLAS W/Z boson rapidity distributions and the corresponding lepton asymmetry data. These data sets were shown at the start of this chapter to not

Observable	W/Z Rapidity	Lepton Asymmetry
MSTW2008	2.01	2.71
MSTWCP	1.63	1.38
MSTWCPdeut	1.56	0.86

Table 5.3.: Table of χ^2 per point (30 (11) points for W/Z (Asymmetry) data) for the standard MSTW2008NLO sets and the new Chebyshev polynomial fits.

agree with the standard MSTW2008 predictions, with particularly poor agreement for the asymmetry. Given the large change in the valence quark distribution for the MSTWCP and MSTWCPdeut fits, and the subsequent improvement in the Tevatron lepton asymmetry agreement, it would be natural to assume an improvement from MSTW2008 will be observed in the ATLAS data.

Table 5.3 demonstrates the improvement in the fit quality for both W/Z rapidity distributions and the lepton asymmetry. Clearly the change in valence quark distribution, and more specifically the $u_v - d_v$ introduced by the new parametrisation fixes the large disagreement for lepton asymmetry. An improvement also occurs for the total W/Z cross section, although the fit is still worse than the CT10 set which remains the best description of this data. The reason for the improvement can be seen in Fig 5.11, where the MSTWCP and MSTWCPdeut bring the theory predictions much more in line with data at low rapidity. This is the region which probes $x \sim 0.01$ for valence quarks, where the extended parametrisation showed the MSTW2008 prediction was most out of agreement.

Finally, the dependence of these cross sections on the eigenvectors of the new fit is studied. As stated previously, there are 23 eigenvector directions in the MSTWCP fit, and the dependence of the W/Z cross section on each of them is shown in Fig 5.12. There is again a significant improvement in the fit in some of the eigenvector directions, most notably number 12, which is most affected by the gluon density. Clearly there is still room for improvement of the central fit

Once again, the predictions for each eigenvector are used in the reweighting procedure to quantify the improvement in fit through eigenvector variations. The results for the $u_v - d_v$ distribution for the two new fits are shown in Fig 5.10. There is still a change in the central value of this distribution required, however the change

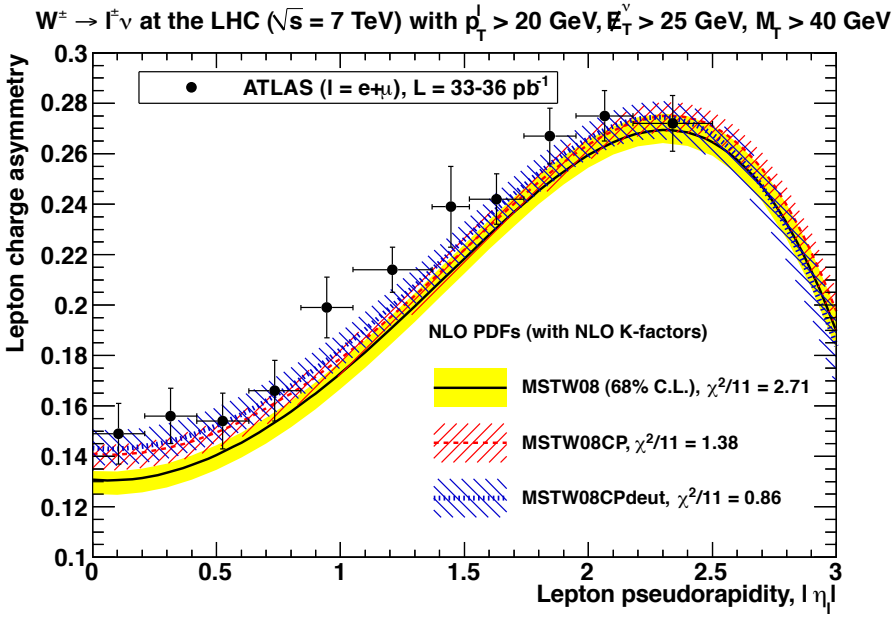


Figure 5.11.: MSTW predictions for ATLAS lepton asymmetry data, with and without the extended parametrisation and deuteron corrections. Plot taken from [82].

in χ^2 between the unweighted and reweighted PDFs is 1.32 to 1.28 per point for the asymmetry data, a 3% improvement. Compared to the 20% improvement in reweighting for the standard MSTW2008 set, it is clear that the new parametrisation provides a better description of this data. To further demonstrate this point, it is notable that the number of effective PDFs in the reweighted set is much higher than for the MSTW2008 set; a sign that the central set is doing better in the new sets.

5.5. Preliminary 2013 PDFs

A new PDF fit has been produced which includes the Chebyshev Polynomial prescription along with the improved deuteron corrections and new theoretical treatment in addition to the inclusion of various updated data sets. The new theoretical treatment is to use the "optimal" GM-VFNS for the treatment of heavy flavour [84], which is smoother near to heavy flavour transition points. Also, the dimuon cross sections have been corrected by the inclusion of a small correction where the charm is produced away from the interaction point. This has a potential impact on the strange quark.

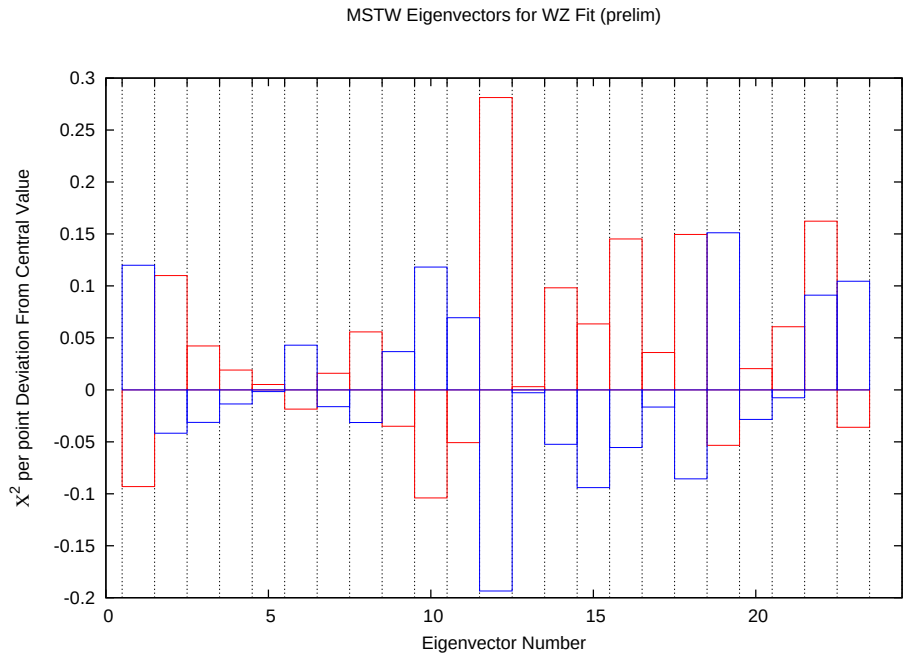


Figure 5.12.: Change in fit quality for each eigenvector in the MSTWCP fit for the ATLAS W/Z total rapidity distributions.

The new data sets include a replacement of the two HERA run I NC data sets with the latest combined H1 and ZEUS data set [85]. The fit to data is excellent for these cross sections. The combined HERA $F_2^{CC}(x, Q^2)$ data [86] is also included, with a fit quality of approximately $60 - 65$ for 52 points. The ZEUS run II NC cross section [87] is additionally included, giving a fit very similar to HERAPDF fits conducted on the data. Finally, all HERA $F_L(x, Q^2)$ measurements [88] [89] [90] are used, and the fit quality obtained is approximately 1 per point.

Some additional and modified Tevatron data is also used in this fit. Firstly, the D0 electron asymmetry data for $p_T > 25$ GeV [91] is newly added along with the CDF W-asymmetry [92]. The χ^2 is approximately 2 per point, and a slight tension is seen between the two data sets. The CDF Z rapidity is changed to the final published numbers [93], which were changed after the MSTW2008 fit was released. Also for this set, the photon contribution is newly added into the theory, although the effect is very small. For this set, no LHC data was included.

These PDF sets are yet to be named, so are referred to as "prelim" in the following discussion. The central value and 68% confidence interval are shown in Fig 5.13, as well as the effect on the ATLAS inclusive jet cross section due to the new parametrisation and data.

The effect moving from MSTW2008 to the prelim PDFs for LHC jet data is minimal. The second plot in Fig 5.13 indicates a change of no more than 2.5%, and the χ^2 fit improves marginally from 0.78 (0.79) for R=0.4 (R=0.6) to 0.74 (0.79). When comparing the CMS inclusive data, which was shown previously to have more relevance for the gluon, a slight decrease in fit quality from 1.48 to 1.50 is observed. Overall, the preliminary PDFs do not seem to have an impact on the LHC jet data, despite giving a significant improvement in the description of the W/Z boson cross sections. This is simply due to the fact that the new parametrisation does not affect the gluon, and so when considering jet data, there is little difference between MSTW2008 and the prelim PDFs. The standard parametrisation used in the MSTW2008 fits can therefore be regarded as sufficient for describing the jet data sets. The effect of the ATLAS inclusive jet cross section on these PDFs is now studied. The plots for both R parameter values are shown in Fig 5.14. The minimal change in the gluon further demonstrates the lack of effect of the new PDF on jet production.

When these new preliminary PDFs are used to analyse the ATLAS W/Z cross section and W asymmetry data, a χ^2 of 1.64 and 0.41 per point respectively was found. The total W/Z fit is then comparable to the CP and CPdeut fits of the previous section, but the asymmetry demonstrates a significantly better fit than

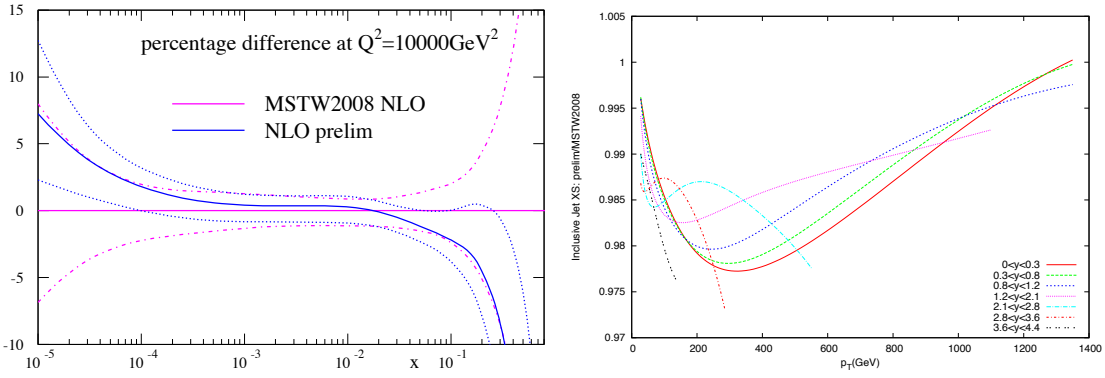


Figure 5.13.: New preliminary PDF gluon compared to the MSTW2008 fit, alongside the change in the ATLAS jet cross section by moving between the two sets.

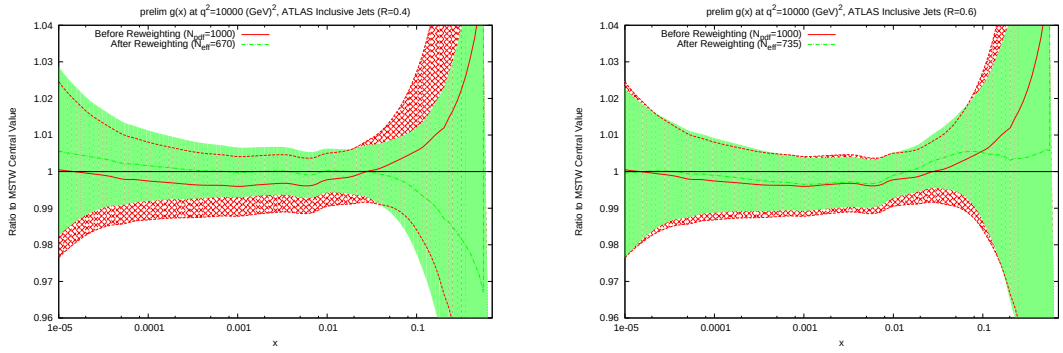


Figure 5.14.: Effect of the ATLAS 7 TeV (left) R=0.4 and (right) R=0.6 inclusive jet data when used to reweight the MSTW prelim PDFs

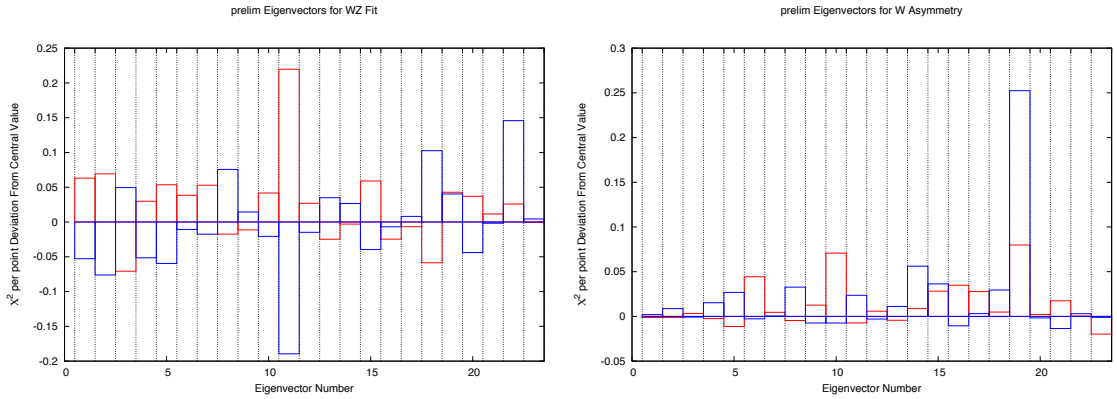


Figure 5.15.: Change in fit quality for each of the 23 eigenvectors in the MSTW prelim fit for (left) W/Z rapidity data and (right) W asymmetry data.

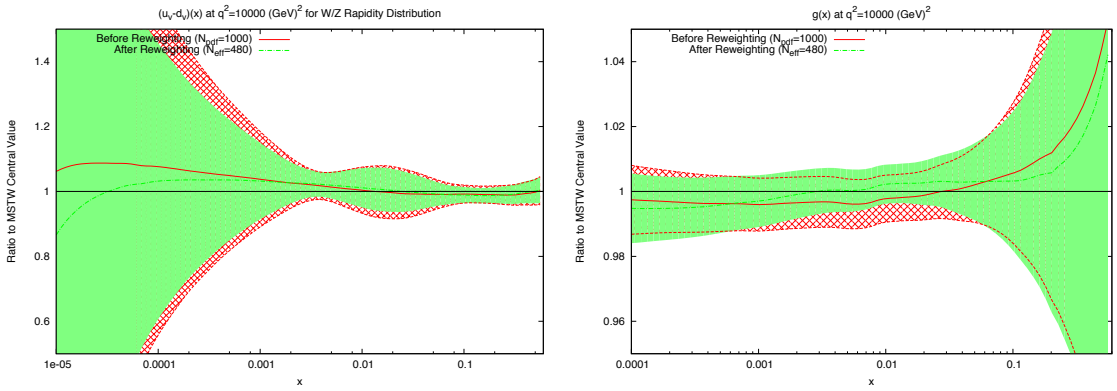


Figure 5.16.: Effect of the W/Z rapidity data when used to reweight the MSTW prelim PDFs.

either of these sets. The potential improvements through the 23 new eigenvectors for the two data sets are shown in Fig 5.15. The quality of fit for the asymmetry data is clear due to the lack of any noticeable reduction in the χ^2 for all eigenvectors. The full rapidity data is most sensitive to eigenvector 11, which is gluon dominated.

The changes in the two most important PDFS, namely $u_v - d_v$ and the gluon, under reweighting to the total W/Z rapidity data is shown in Fig 5.16. There is little change in the central value for the valence difference, with only a reduction in the error bands occuring. In comparison, the gluon central value shifts upwards for moderate values of x, and so clearly has the larger impact on the data. The fit quality after reweighting is 1.44 per point, a reasonable improvement.

5.6. Summary

In this section, the W and Z boson rapidity distributions at ATLAS have been studied from the point of view of MSTW PDFs. These distributions are shown to not be in good agreement with the standard MSTW2008 set compared to the other major PDF groups. The origin of this disagreement is the $u_v - d_v$ distribution, which is not properly described by the current parametrisation used my MSTW. The intermediate x region, which is affected most by the polynomial term in the parametrisation, is shown to not provide enough flexibility in the $u_v - d_v$ distribution. A new method based on the use of Chebyshev Polynomials is shown to improve the χ^2 fit for these data sets, and along with a new treatment of deuteron corrections,

the fit is brought in line with the other PDF groups. A new preliminary PDF fit is tested which includes these changes along with new HERA data. This new set is shown to improve the ATLAS jet fit marginally, and worsen the CMS fit by an equally small amount. Neither effect is significant, since the new parametrisation does not affect the gluon, which is the dominant factor in the jet cross section.

Chapter 6.

New PDF Sets

In this chapter, new PDF sets are produced including the LHC inclusive jet data discussed in the previous chapters. There was sufficient motivation from the eigenvector reweighting studies into the ATLAS combined 2.76 TeV and 7 TeV data and the CMS data to justify a new set. In addition, this is an opportunity to further test the validity of the reweighting technique as a general method of quantifying the effect of a new data set on PDFs. Two fits are performed in this chapter, the first of which includes only the ATLAS 7 TeV and CMS inclusive data, both of which were calculated with FastNLO version 2. The second fit additionally includes the ATLAS combined data, which is calculated using APPLgrid and required modifications to the fitting code.

6.1. Fit With ATLAS 7 TeV and CMS Inclusive Jet Data

In order to include the CMS data into an MSTW fit, the first necessary task was to modify the fit code to include FastNLO version 2 [94]. This new version allows more scale flexibility within the cross section calculation, however the interface required modification to the α_s and PDF convolution routines. The MSTW fitting code was therefore modified to utilise the new version, and can now read all future LHC FastNLO routines as they become available.

The fit is performed allowing the same parameters to be free as in the standard MSTW2008 set. Initially, α_s was allowed to be free, and a reasonable improvement

in the global fit from 2795 to 2781 over 2966 data points was obtained. This fit, however, included a decrease in α_s from 0.1202 to 0.1189, which caused much of the improvement. Subsequently, in order to properly quantify the effect on just the PDFs, α_s was held fixed at 0.1202. This fit yielded a smaller improvement of only 8 points from 2795 to 2786. These two new fits are labelled in this thesis as MSTWCMS α_s Free and MSTWCMS α_s Fixed.

The effect on each data set included in the fit is shown in Table 6.1. The ATLAS and CMS χ^2 values for MSTW2008 were first calculated using the fitting code by passing through the central value and bypassing the minimisation steps. Once they are included in the minimisation, a very large improvement in the fit to CMS data is seen with a more modest improvement for the ATLAS data. The fact that both data sets prefer a smaller α_s is shown in the fact that the improvement is less pronounced when it is held fixed. In general, the fit to the various DIS data sets is left unchanged by both of the new fits. The Tevatron inclusive jet fits worsen very slightly with the inclusion of the LHC data, although on the whole the Tevatron data remains also unchanged. The improvement of the global fit with α_s free can be understood through the stark improvement in the BCDMS proton F_2 measurement. This set returns to its original χ^2 value once α_s is fixed. One data set which has a noticeable difference between the α_s convention is the NuTeV xF_3 measurement. This is a non-singlet distribution, and so its evolution is driven entirely by α_s. Ay change in α_s will then affect this data set, and the choice of α_s that provides the best fit to the global data is seen to provide a worse fit for this data set. These PDFs will be named here MSTWCMS, due to the dominance of the CMS inclusive jet data on the improvement in fit quality.

The new central PDF is shown in Fig 6.1, along with the reweighted PDF using the CMS inclusive data. The two error bands shown are the original MSTW2008 68% confidence level, and the reweighted standard deviation of the randomly generated PDFs. It is clear that the new PDF requires a similar behaviour in the gluon as the reweighting technique. Whilst the two central lines do not exactly match, there is a trend for a $\sim$1% increase in the gluon for much of the x range, which turns into a rapidly decreasing gluon at around $x \sim 0.1$. The error band of the reweighted PDF is in good agreement with that of the new fit for most values of x. The only region with disagreement is at high-x, where the reweighting technique appears to

Data Set	MSTW2008	MSTWCMS α_s Free	MSTWCMS α_s Fixed
BCDMS μp F_2	182/163	172/163	182/163
BCDMS μd F_2	190/151	188/151	189/151
NMC μp F_2	121/123	122/123	120/123
NMC μd F_2	102/123	103/123	102/123
NMC $\mu p/\mu d$	130/148	131/148	130/148
E665 μp F_2	57/53	54/53	54/53
E665 μd F_2	53/53	57/53	57/53
SLAC μp F_2	30/37	30/37	30/37
SLAC μd F_2	30/38	33/38	30/38
NMC/BCDMS/SLAC F_L	38/41	40/31	38/31
E866/NuSea pp DY	228/184	227/184	229/184
E866/NuSea pd/pp DY	14/15	13/15	14/15
NuTeV νN F_2	49/53	50/53	50/53
CHORUS νN F_2	26/42	26/42	26/42
NuTev νN xF_3	40/45	45/45	40/45
CHORUS νN xF_3	31/33	32/33	31/33
CCFFR $\nu N \to \mu\mu X$	66/86	66/86	65/86
NuTeV $\nu N \to \mu\mu X$	39/40	39/40	40/40
H1 MB 99 e^+p NC	9/8	9/8	9/8
H1 MB 97 e^+p NC	42/64	43/64	44/64
H1 low Q^2 96-97 e^+p NC	44/80	44/80	45/80
H1 high Q^2 98-99 e^-p NC	122/126	122/126	120/126
H1 high Q^2 99-00 e^+p NC	131/147	131/147	128/147
ZEUS SVX 95 e^+p NC	35/30	35/30	35/30
ZEUS 96-97 e^+p NC	86/144	86/144	85/144
ZEUS 98-99 e^-p NC	54/92	54/92	53/92
ZEUS 99-00 e^+p NC	63/90	63/90	62/90
H1 99-00 e^+p CC	29/28	29/28	29/28
ZEUS 99-00 e^+p CC	38/30	38/30	38/30
H1/ZEUS ep F_2^{charm}	107/83	106/83	109/83
H1 99-00 e^+p incl. jets	19/24	17/24	18/24
ZEUS 96-97 e^+p incl. jets	30/30	29/30	29/30
ZEUS 98-00 $e^\pm p$ incl. jets	17/30	16/30	16/30
D0 II $p\bar{p}$ incl. jets	114/110	116/110	115/110
CDF II $p\bar{p}$ incl. jets	56/76	60/76	58/76
CDF II $W \to l\nu$ asym.	29/22	30/22	29/22
D0 II $W \to l\nu$ asym.	25/10	28/10	26/10
D0 II Z rap.	19/28	17/28	19/28
CDF II Z rap.	49/29	50/29	50/29
ATLAS 7TeV incl. jets (R=0.4)	(72/90)	66/90	70/90
CMS 7 TeV incl. jets	(180/133)	163/133	169/133
Total	2795/2922	2781/2922	2786/2922

Table 6.1.: Table of χ^2 values for each data set included in the fits for the standard MSTW 2008 NLO fit and the new NLO fits with ATLAS 7 TeV and CMS data. The ATLAS and CMS values are quoted for MSTW 2008 despite not being included in the fit. These are simply the χ^2 values obtained when the fit code is run using the standard set without minimisation.

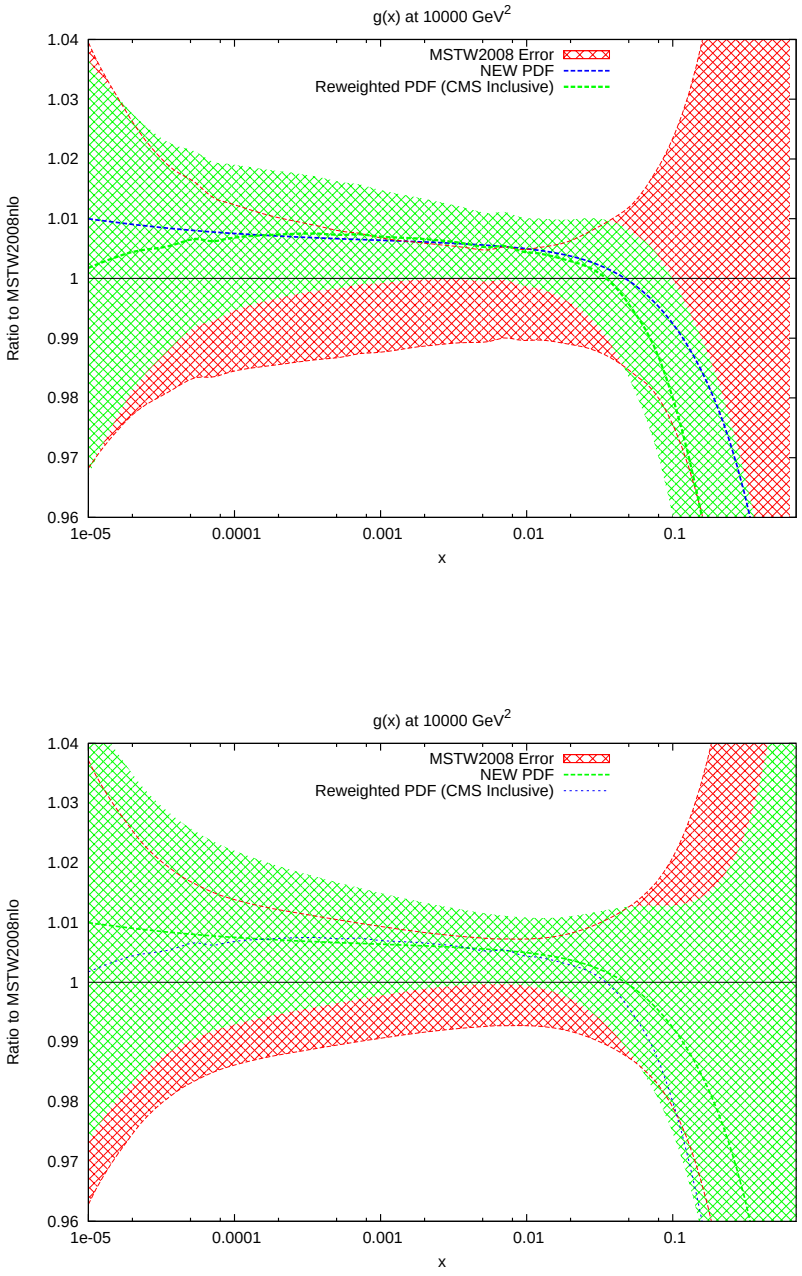


Figure 6.1.: Comparison of the gluon for standard MSTW fit, reweighted PDF (using CMS inclusive jets to reweight), and the new fit directly including the ATLAS & CMS data. All 3 central values are shown on each plot; the first compares the error bands for MSTW against reweighting, and the second compares standard MSTW to the new fit.

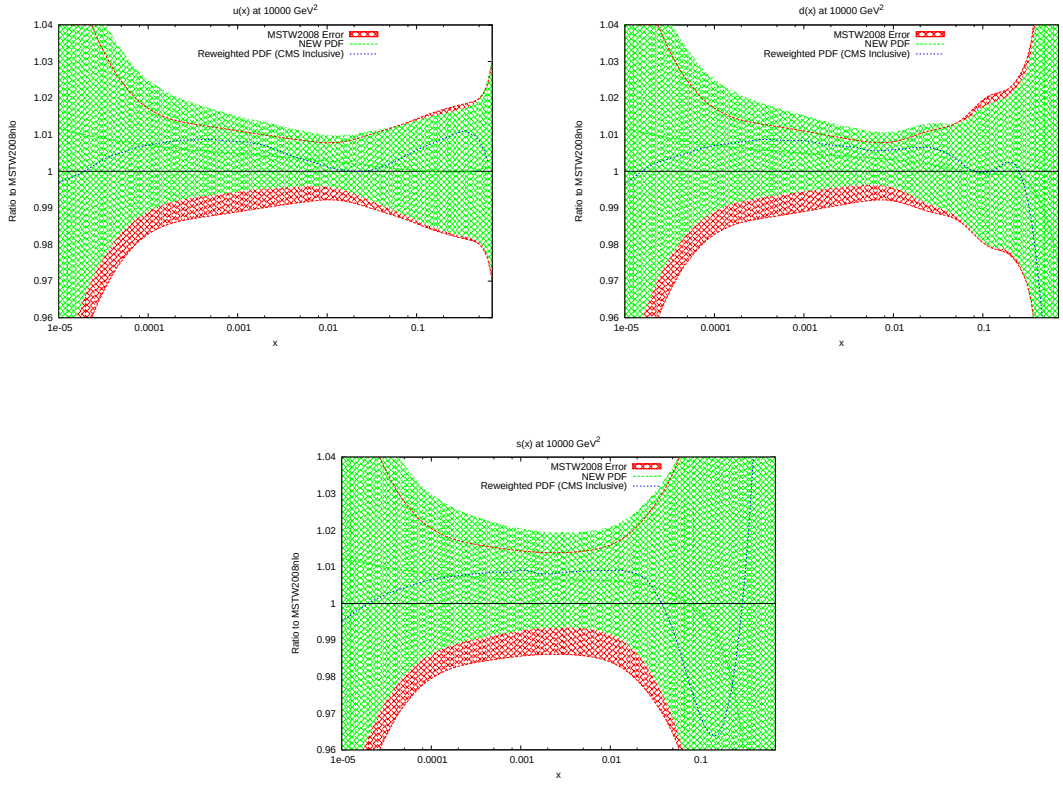


Figure 6.2.: Ratio of the MSTWCMS quark distributions to MSTW2008. The central value of the reweighted PDF using CMS inclusive data is also shown for comparison.

underestimate the error. Upon inspection of the top weighted PDFs used, all require a steeply falling gluon compared to MSTW2008, and so the standard deviation shows a strong grouping around this trend.

The new quark PDFs are shown in Fig 6.2. These were shown to be important for the CMS inclusive jet data in Chapter 3 due to the probed x values and resulting partons probed. The magnitude of change from MSTW2008 is similar to the gluon for all of these distributions, lending further evidence for the importance of these PDFs. Again, there is good agreement between the reweighting technique and the direct inclusion of data. The only large disagreement is in the high-x strange distribution where the uncertainties are very large. The absolute size of the PDF uncertainties are shown in Fig 6.3. Here we can see that MSTWCMS has a slightly better constrained

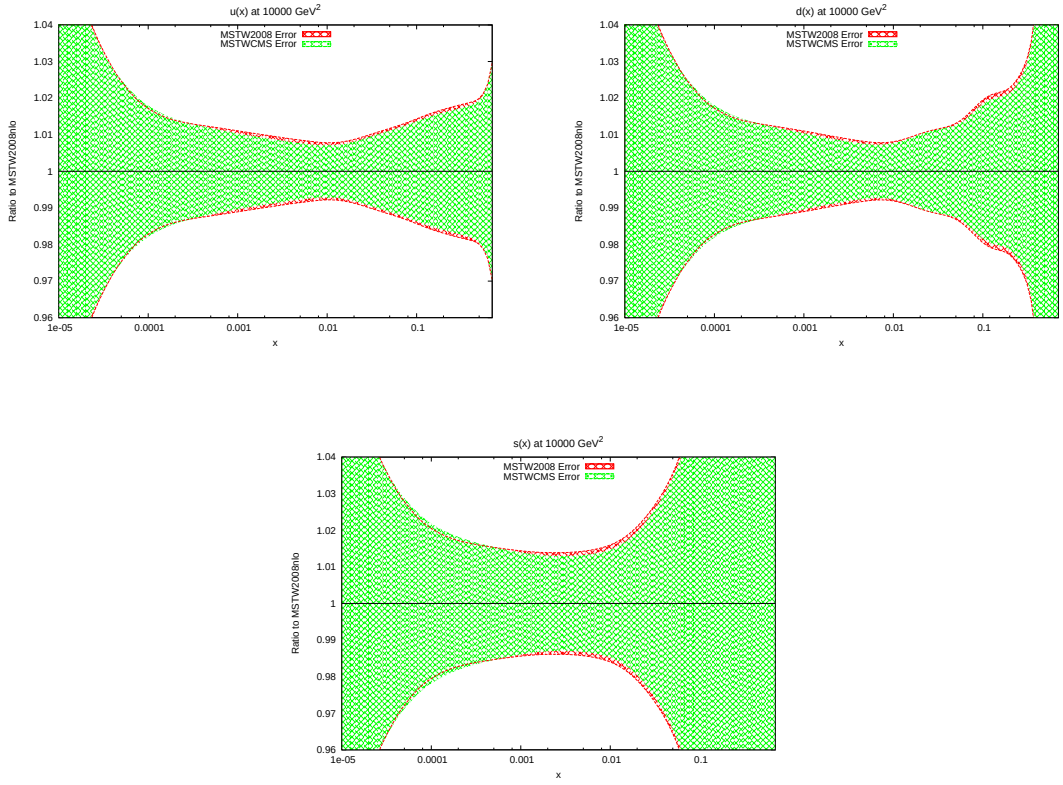


Figure 6.3.: Absolute size of PDF uncertainties for MSTW2008 and MSTWCMS relative to the respective central set. This figure demonstrates the slight improvement in PDF constraint for the quark PDFs when including the CMS data in the global fit.

set of quark PDFs, with the error bands reducing in size for all but the lowest x values.

The new prediction for CMS inclusive jets is shown in Fig 6.4. There is no change in shape between the new prediction and the MSTW 2008 prediction, and most points lie within the experimental error bars. However a systematic downward shift of $\sim 1\%$ is seen across most data points. For lower rapidity bins, where the experimental error is smallest, this shift brings some points out of agreement with the MSTW prediction, and this is where the largest change in χ^2 originates.

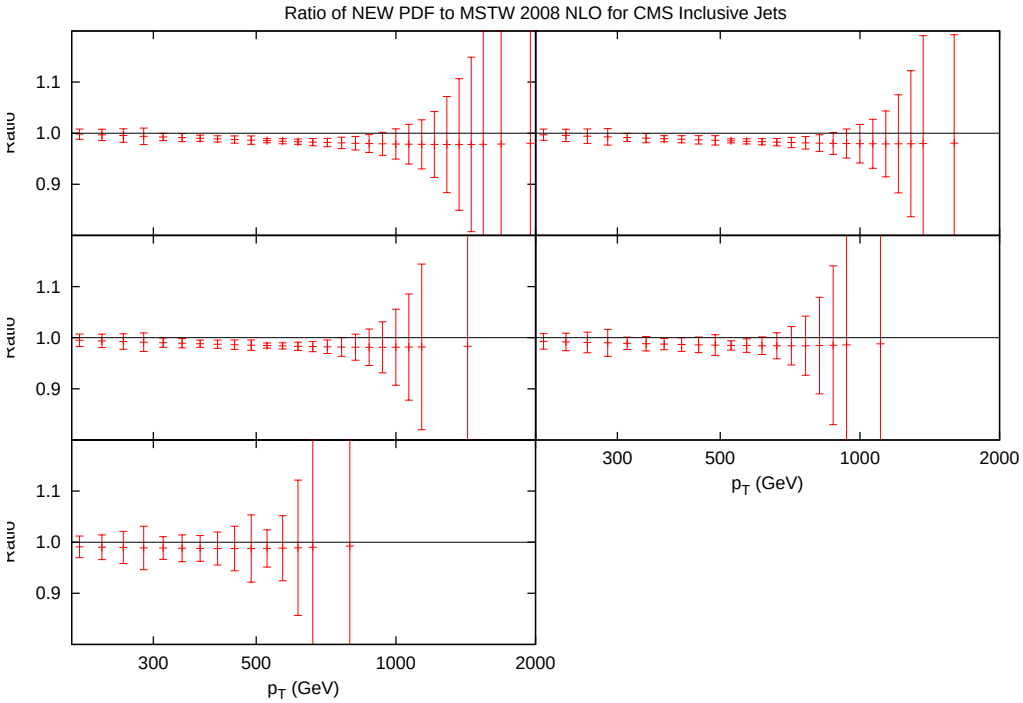


Figure 6.4.: Ratio of CMS inclusive jet cross section predictions for the new PDFs and the standard MSTW 2008 PDFs.

6.2. Eigenvectors

The eigenvectors for the new fit are calculated in the same manner as the usual MSTW global fits. There are again 20 eigenvectors due to the same parameters being free, however the dependence of each eigenvector on the underlying parameters and data sets has changed. The fractional contribution to the total uncertainty on the major distributions from each eigenvector is shown in Figs 6.5 - 6.14. These can be interpreted as the sensitivity to the underlying PDFs of each eigenvector, and can be compared to the equivalent plots for the MSTW2008 fit presented in [4]. These plots can be used to infer which distributions are most important to a data set given the sensitivity of the data set to each eigenvector. For example, the u_v distribution at $x \sim 0.1$ can be seen to be almost entirely governed by the new eigenvector 18. Data sets which probe this particular region of this valence up quark distribution will be highly influenced by this eigenvector. When considering a different distribution, such as the low x gluon, the contribution to the error can be seen to be spread between

eigenvectors 1, 9, 11 and 19. Any or all of these could then have an effect on inclusive jet data, which is most sensitive to these distributions.

The shapes present in these plots are highly antisymmetric in the direction of the uncertainty, as can be seen prominently in the $xs - x\bar{s}$ distribution for eigenvector 6. This is due to the use of the antisymmetric error definition used in the calculation of the PDF errors. Each eigenvector will affect the distributions differently when moved away from the global minimum in opposing directions.

The CMS data itself directly constrains eigenvector 19 in this set. This can be seen to be almost entirely dependent on the gluon, although the up valence quark is also affected. Both distributions are most sensitive to this eigenvector in the high x region, which is consistent with the conclusions of the reweighting study, where the gluon and quark distributions were shifted the most in this region after reweighting to the CMS data.

The change in fit quality to the ATLAS inclusive jet combined data for each of the new eigenvectors is shown in Fig 6.15 alongside the corresponding plot for MSTW2008. There is more dependence on the eigenvectors of the MSTW2008 set, and large increases in χ^2 can be obtained for many eigenvectors. The new eigenvectors do not produce this dramatic reduction in fit quality, implying a better agreement with the data. Despite this, there are still many eigenvectors which can improve the fit to a reasonable degree. The largest are eigenvectors 2, 6 and 14. The new PDF can then be said to provide a better fit to ATLAS combined data, with scope for further improvement.

6.3. Reweighting of the New PDFs

An important study which can now be performed is to reweight the new PDFs, which will again check the compatibility of the method with the standard fitting procedure. By using the new central value and eigenvectors, the χ^2 for ATLAS combined jets is calculated for 1000 PDFs randomly generated in the eigenvector space. The distribution can then be compared to that of the PDFs randomly distributed in the standard MSTW eigenvector space.

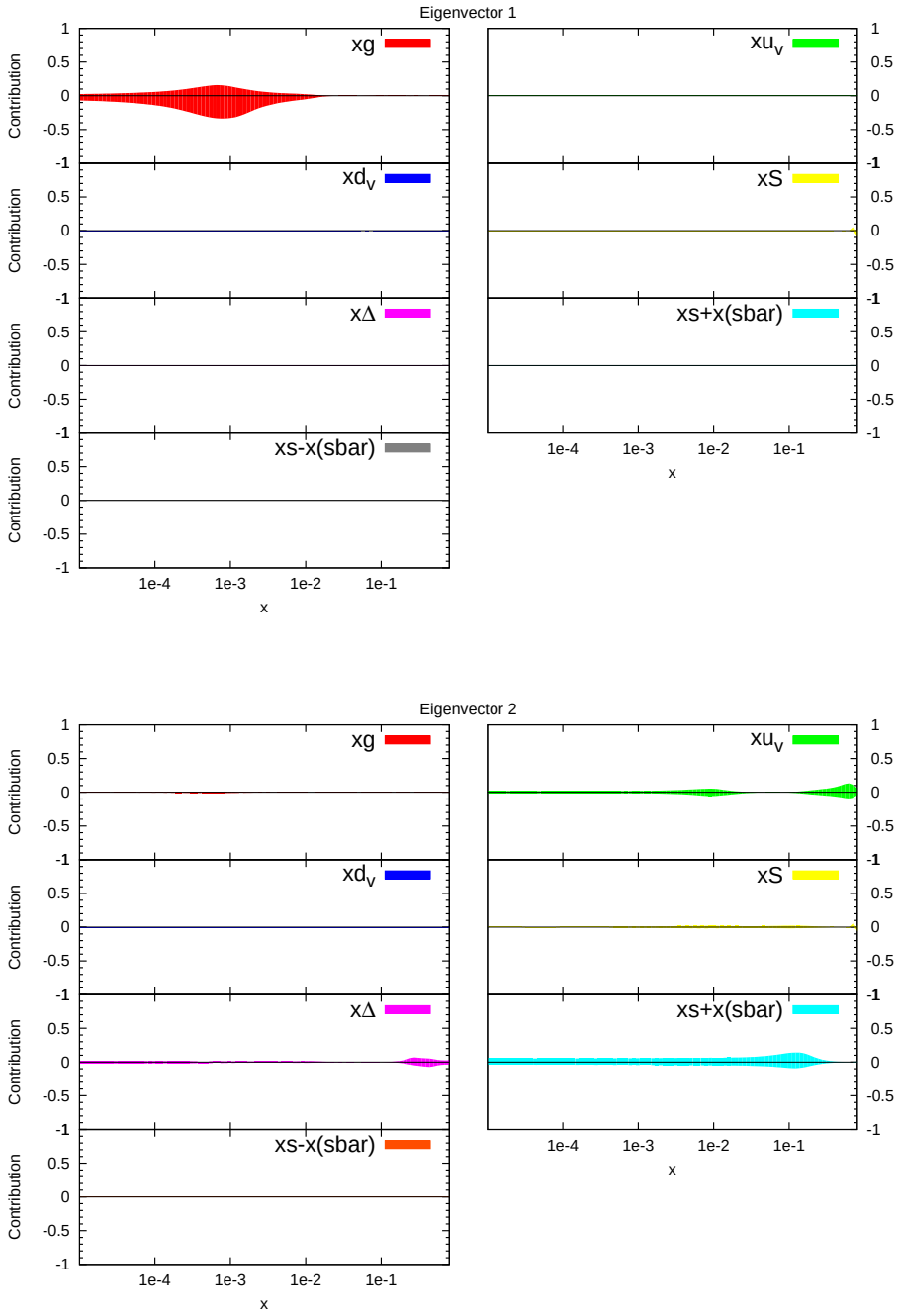


Figure 6.5.: Fractional contribution to the uncertainty on major distributions from each eigenvector. Eigenvector 1 and 2 shown.

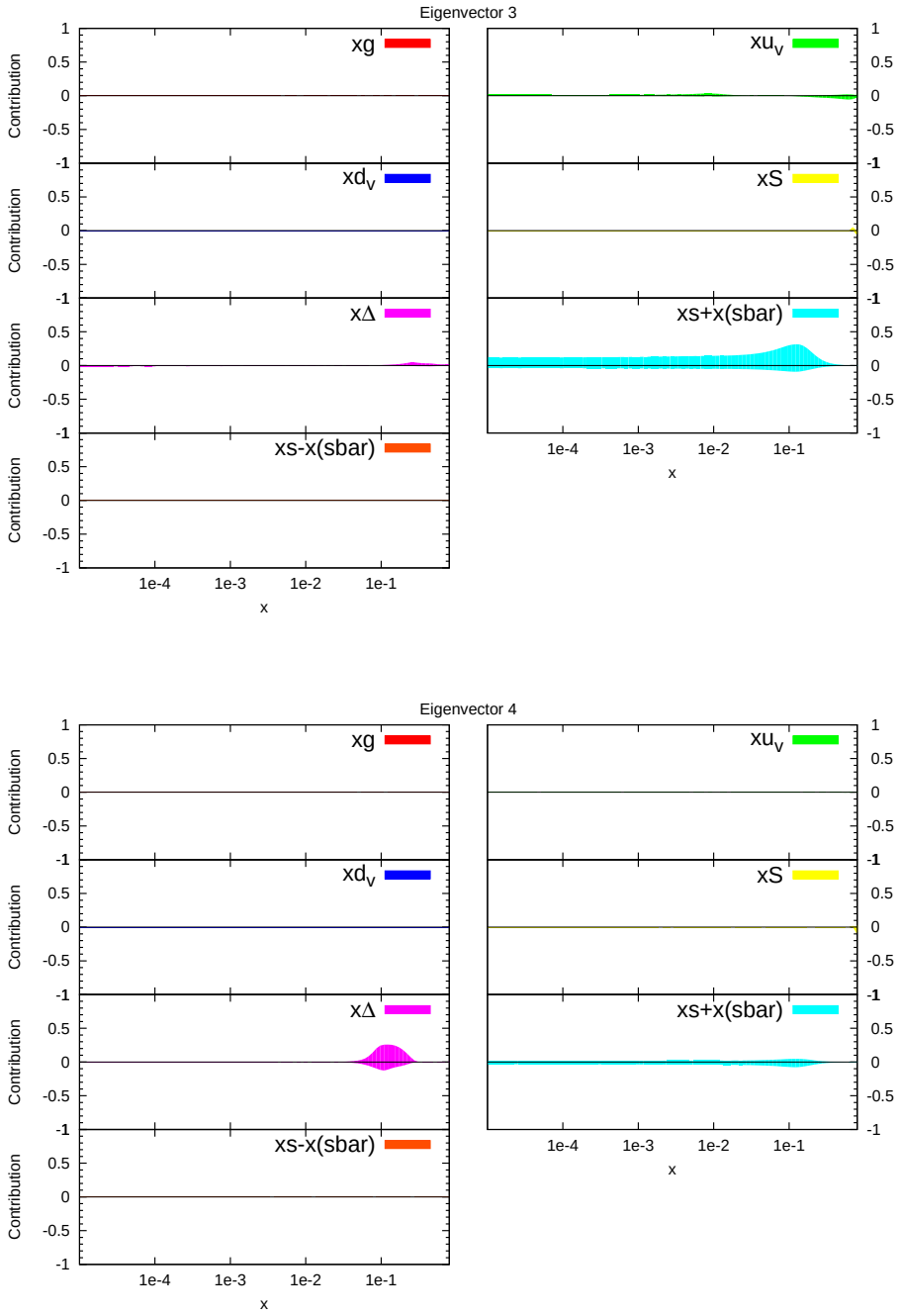


Figure 6.6.: Fractional contribution to the uncertainty on major distributions from each eigenvector. Eigenvector 3 and 4 shown.

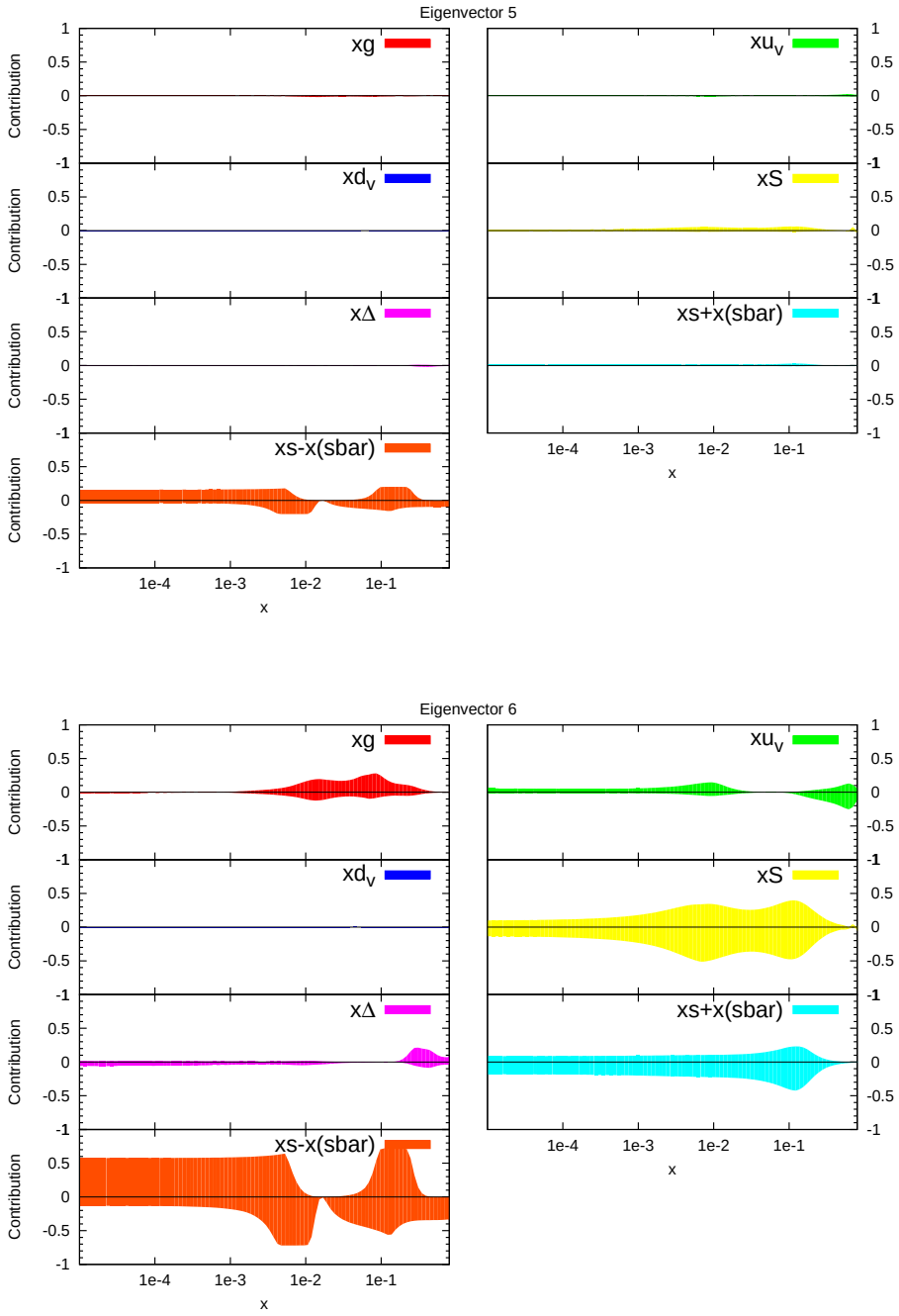


Figure 6.7.: Fractional contribution to the uncertainty on major distributions from each eigenvector. Eigenvector 5 and 6 shown.

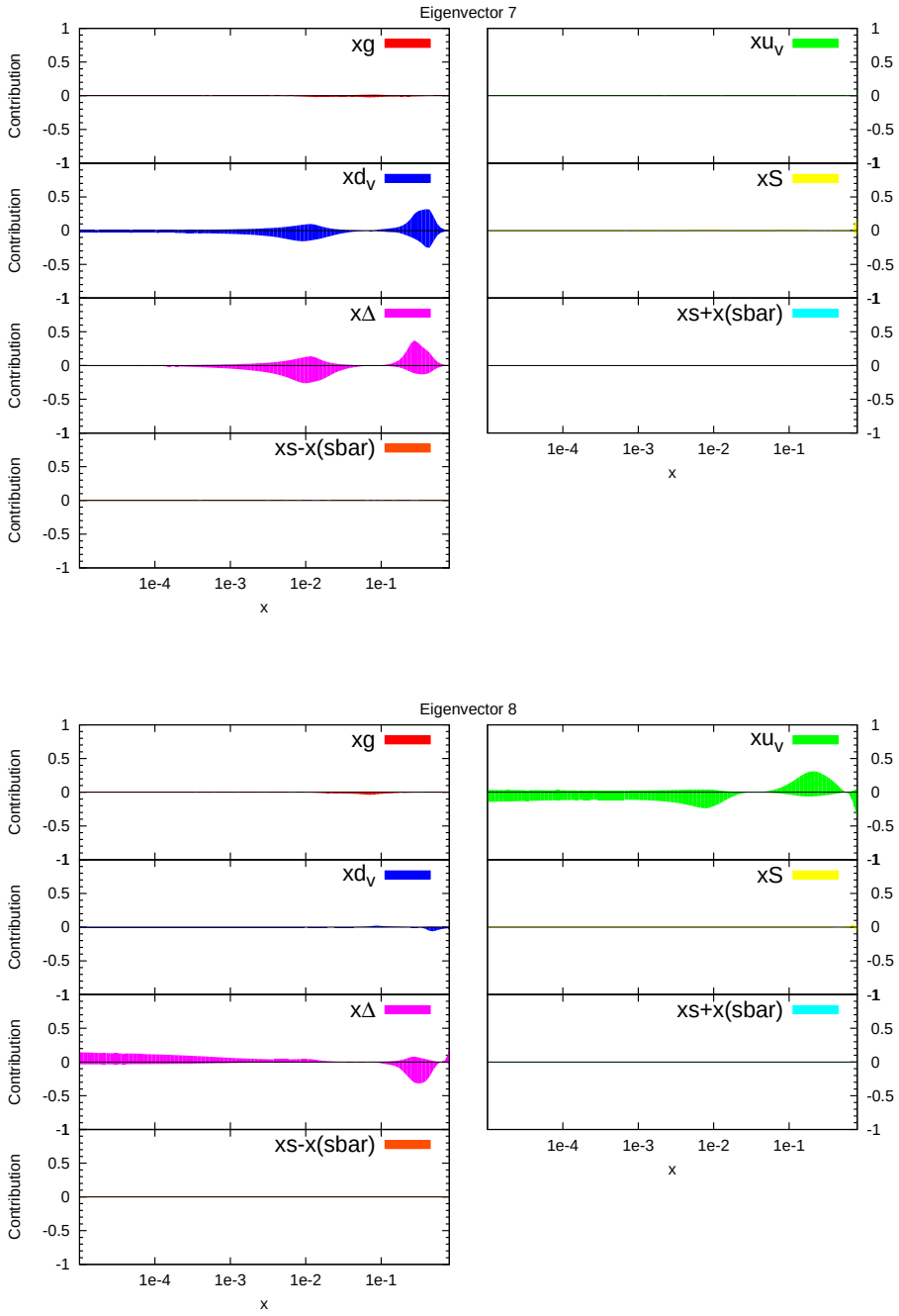


Figure 6.8.: Fractional contribution to the uncertainty on major distributions from each eigenvector. Eigenvector 7 and 8 shown.

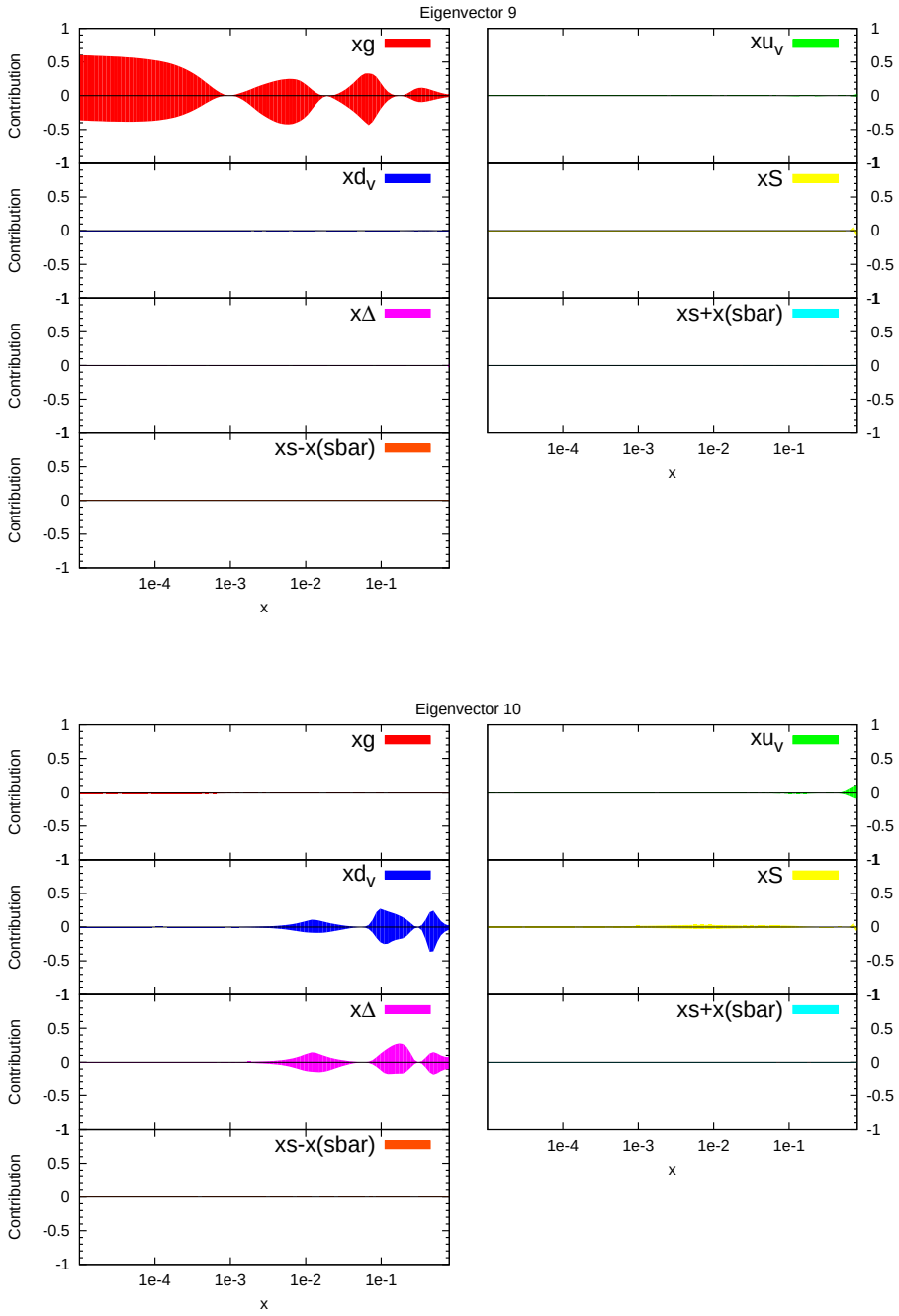


Figure 6.9.: Fractional contribution to the uncertainty on major distributions from each eigenvector. Eigenvector 9 and 10 shown.

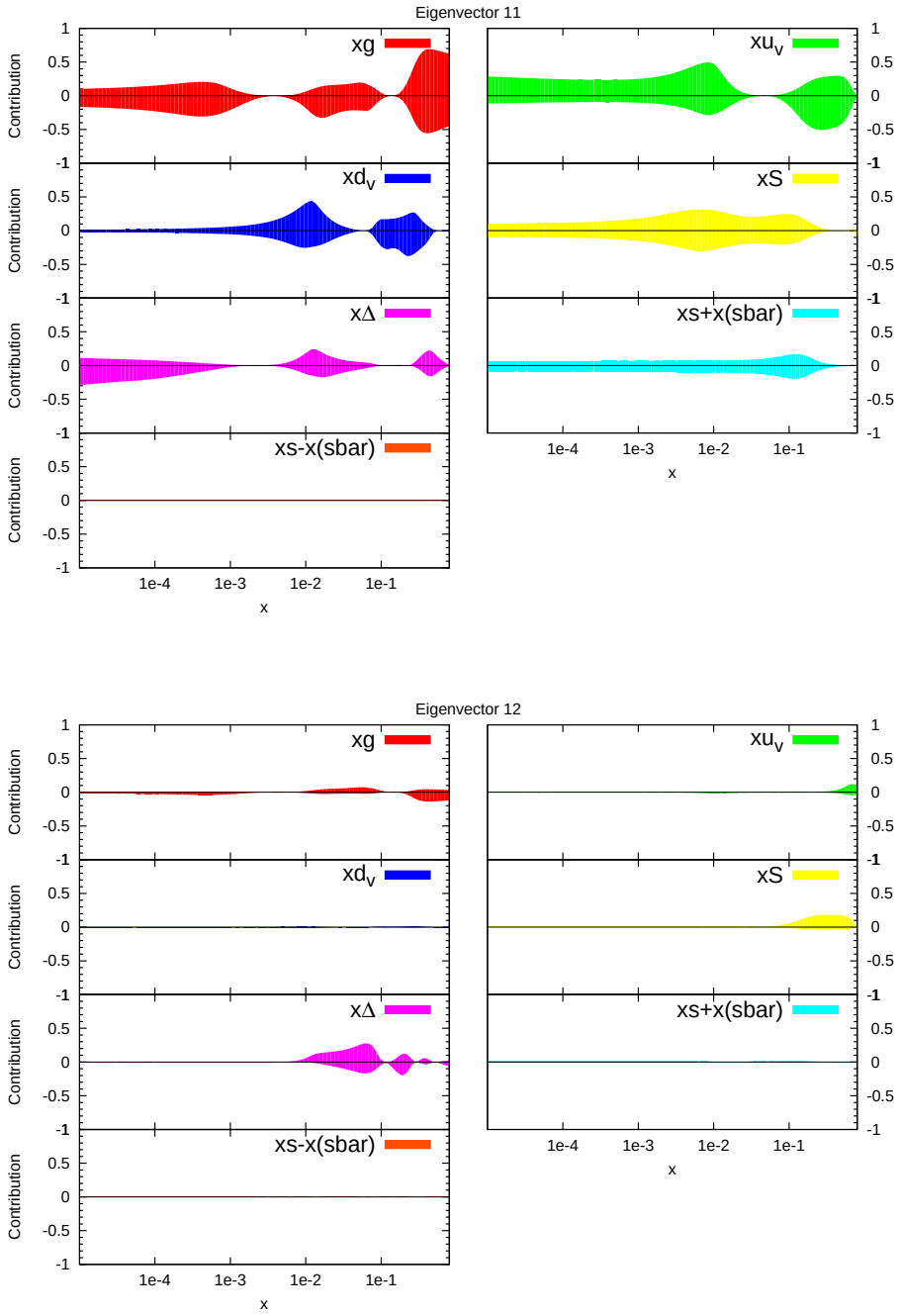


Figure 6.10.: Fractional contribution to the uncertainty on major distributions from each eigenvector. Eigenvector 11 and 12 shown.

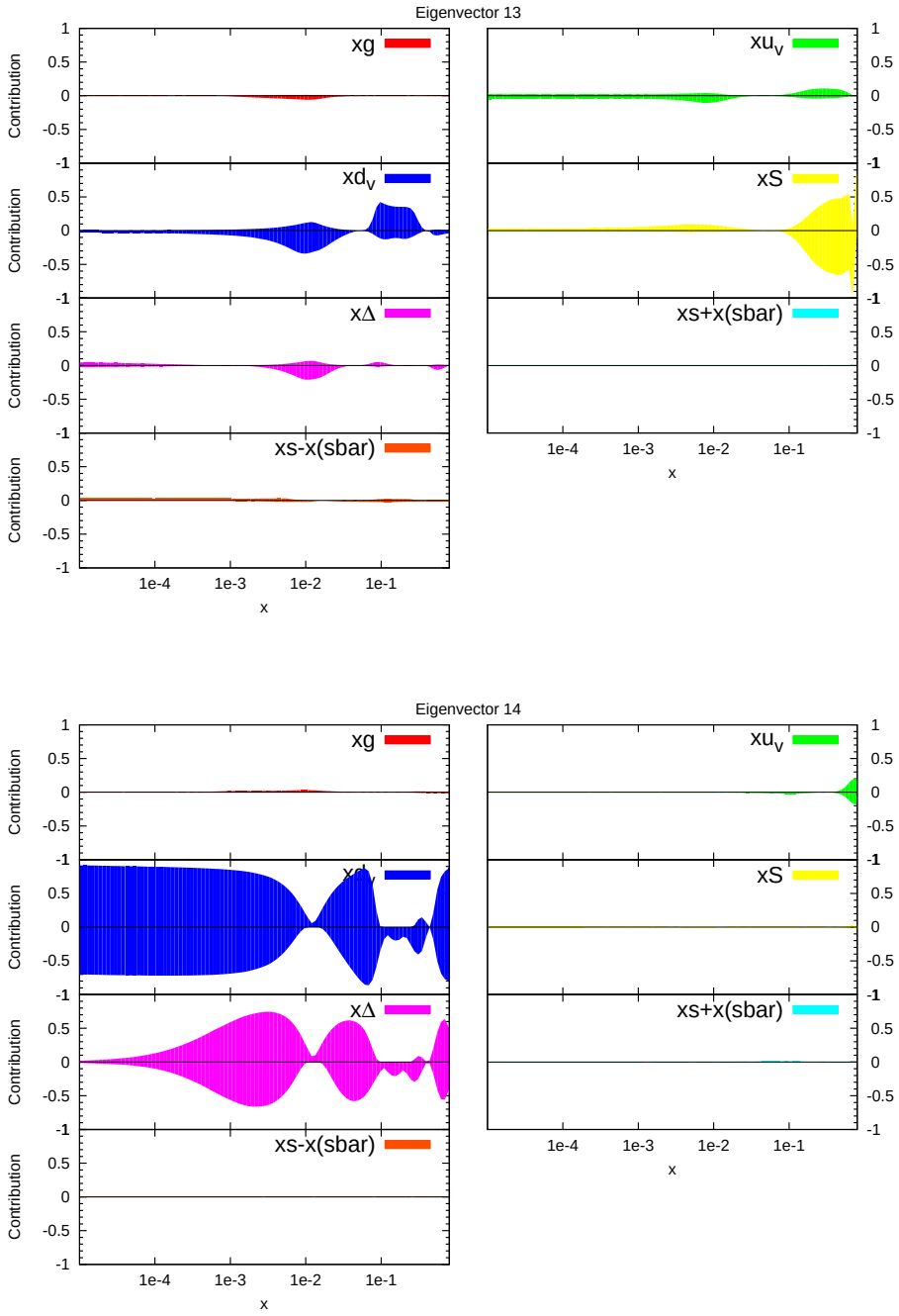


Figure 6.11.: Fractional contribution to the uncertainty on major distributions from each eigenvector. Eigenvector 13 and 14 shown.

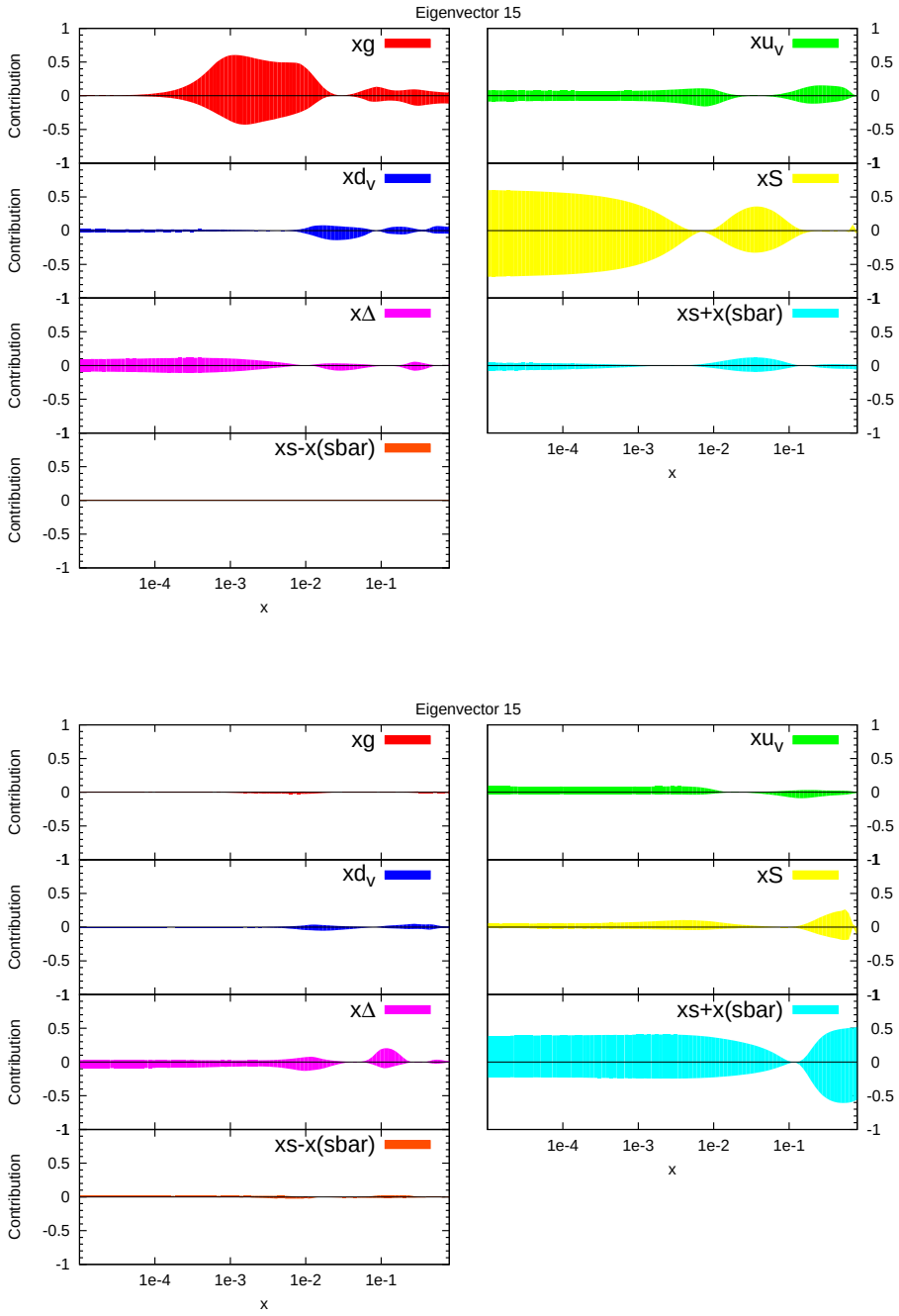


Figure 6.12.: Fractional contribution to the uncertainty on major distributions from each eigenvector. Eigenvector 15 and 16 shown.

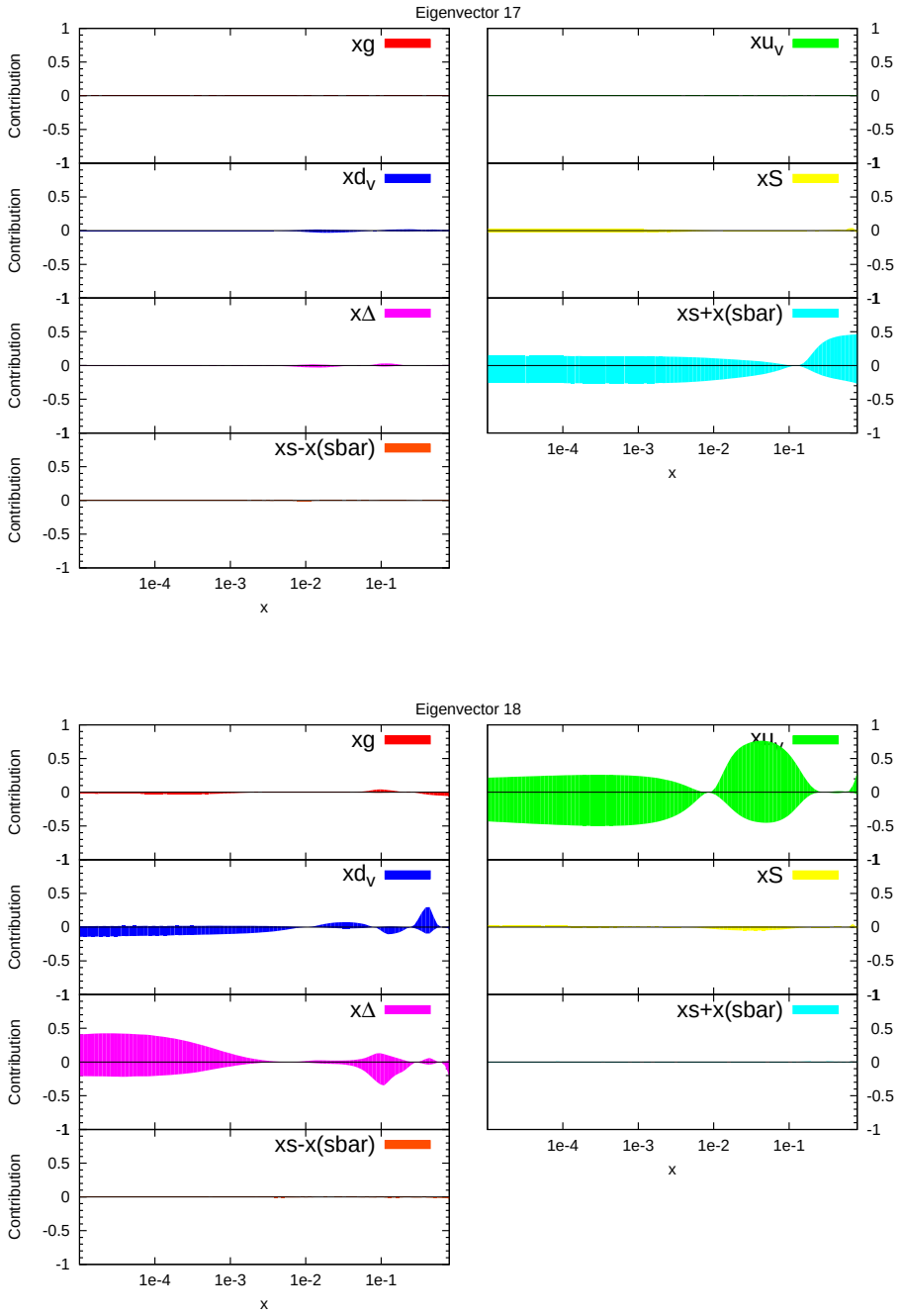


Figure 6.13.: Fractional contribution to the uncertainty on major distributions from each eigenvector. Eigenvector 17 and 18 shown.

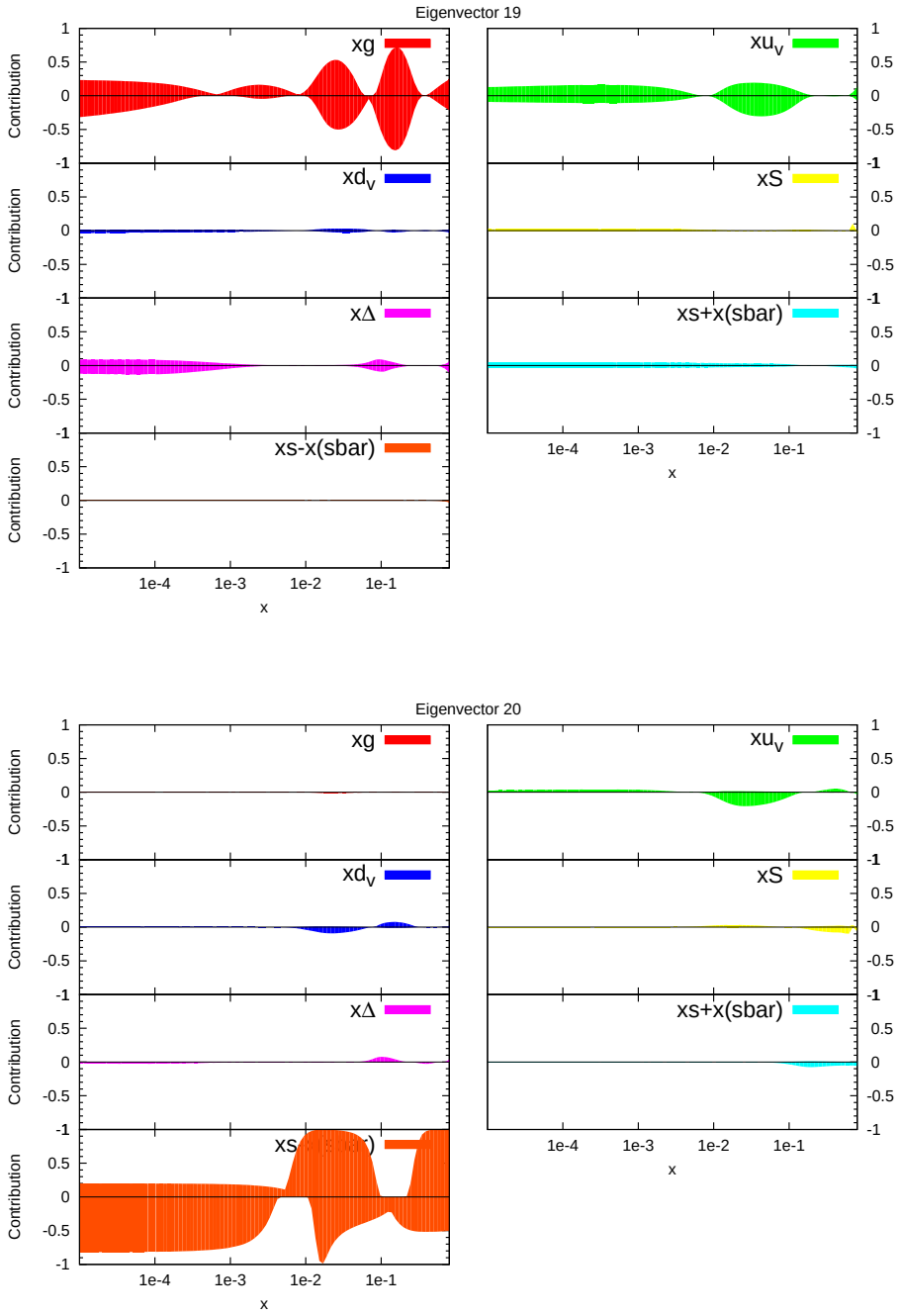


Figure 6.14.: Fractional contribution to the uncertainty on major distributions from each eigenvector. Eigenvector 19 and 20 shown.

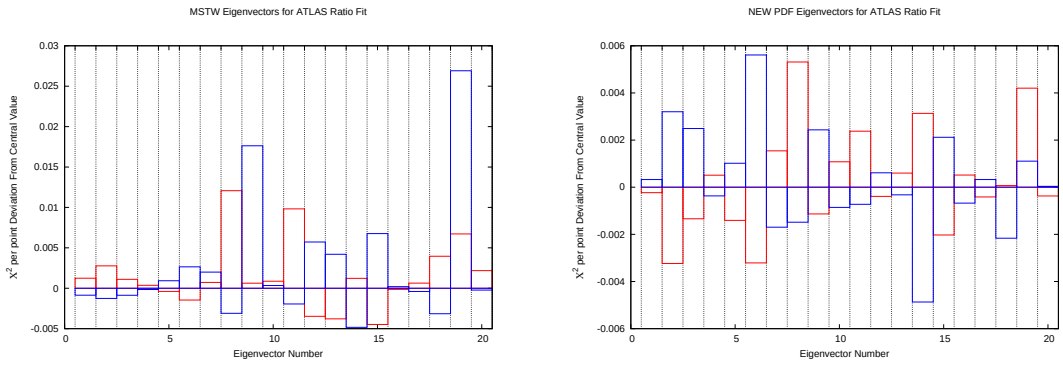


Figure 6.15.: Change in fit quality to the ATLAS combined 2.76 TeV and 7 TeV cross sections from the MSTW2008 (left) and MSTWCMS (right) central values for each eigenvector in the respective fits.

The ATLAS combined 2.76 TeV and 7 TeV inclusive jets were not initially included in the fit due to the fact that the predictions were only available for APPLgrid and not FastNLO. The data is partially included in the form of the 7 TeV cross section only, and so the new PDFs are expected to somewhat replicate the effect of this data set. The observed effect is shown in Fig 6.16. The cuts discussed in Chapter 3 are used along with the additive treatment. There is still a shift required of the gluon under reweighting, and the effect is in fact slightly larger than when this data set was reweighted in Chapter 3. This can be interpreted as further evidence that the 7 TeV ATLAS inclusive data has little effect on the PDFs, and the combined data including the 2.76 TeV set must be used.

Finally the dijet cross sections are studied using the new PDFs. After the previous studies which showed that in general the dijet data sets require a different shift in the PDFs to the equivalent inclusive jet data, this is the ideal test of compatibility between the data types.

Fig 6.17 demonstrates the effect of the CMS dijet data on the new PDFs. There is in fact very little difference between the shape of the reweighted gluon with respect to the new PDF as that with respect to the MSTW central value. In fact, the slight reduction in the error band for the new PDF causes the reweighted central value to be outside of the error band for a small x range. The trend is still opposing the inclusive jet data, with a smaller gluon required at moderate x, and a larger gluon at low x. The reweighted PDF has a χ^2 of 1.77 per point, compared to the unweighted

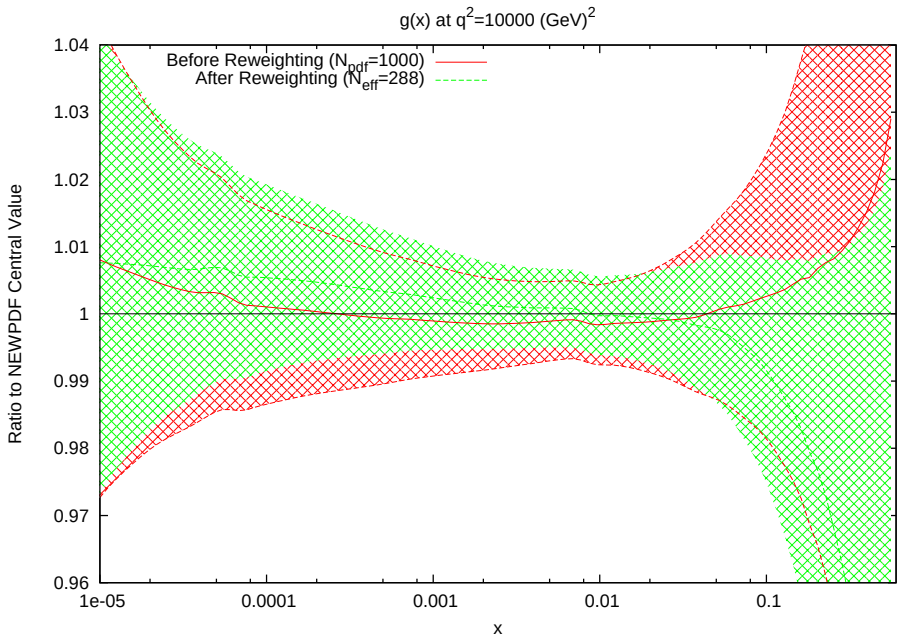


Figure 6.16.: Reweighting of the new gluon PDF using ATLAS combined data.

central value which is 2.02 per point. Both of these values are larger than the 1.67 per point which is the value after reweighting to the MSTW 2008 PDFs, which implies that the new PDFs are in fact worse at describing the CMS dijet data, despite the corresponding inclusive jet data being newly included in these sets.

6.4. $\Delta\chi^2 = 1$ Treatment

Until now, the reweighting procedure used in this thesis and in the previous MSTW study have used the standard MSTW eigenvectors, which are defined using dynamical tolerance levels. In this procedure, some eigenvectors are allowed to move further from the global minimum in χ^2 than others, depending on the functional behaviour about the minimum in the relevant direction. However, when reweighting using the eigenvectors, it may be more appropriate to instead use a set tolerance of $\Delta\chi^2 = 1$ in each direction.

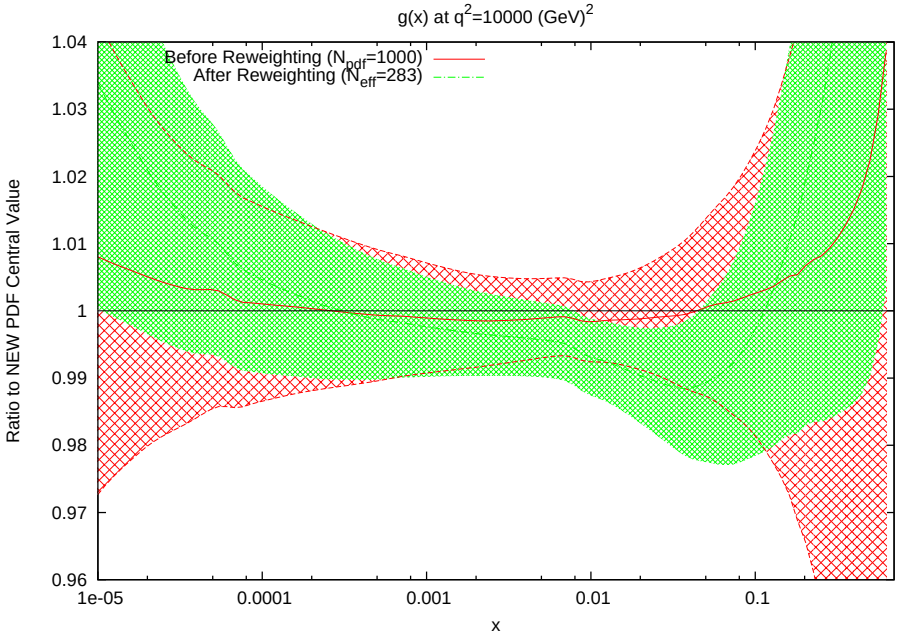


Figure 6.17.: Reweighting of the new gluon PDF using CMS dijet data (p_T^{av} scale choice)

The reweighted gluon using this technique is shown with the gluon of the new fit PDF in Fig 6.18. Whilst the reweighting had previously agreed well with the required shift for the new PDF, there is clear disagreement here. The new PDF is well outside the 1σ error band. This can be explained simply by an inability for the the random PDFs to be generated in the required range. Given that on average, the dynamic tolerance level for the eigenvectors in the MSTW2008 fit are approximately 3 to 4, by rescaling to a value of 1, we can assume that all error bands will be reduced by a factor of 3 or 4.

The 1000 randomly distributed PDFs using the $\Delta\chi^2 = 1$ method are shown in Fig 6.19, and demonstrate this inability to replicate the new PDF gluon. Whilst a very small handful extend to the required upward shift, these are drowned out by the vast majority which, whilst weighted lightly, contribute the most to the reweighted PDF.

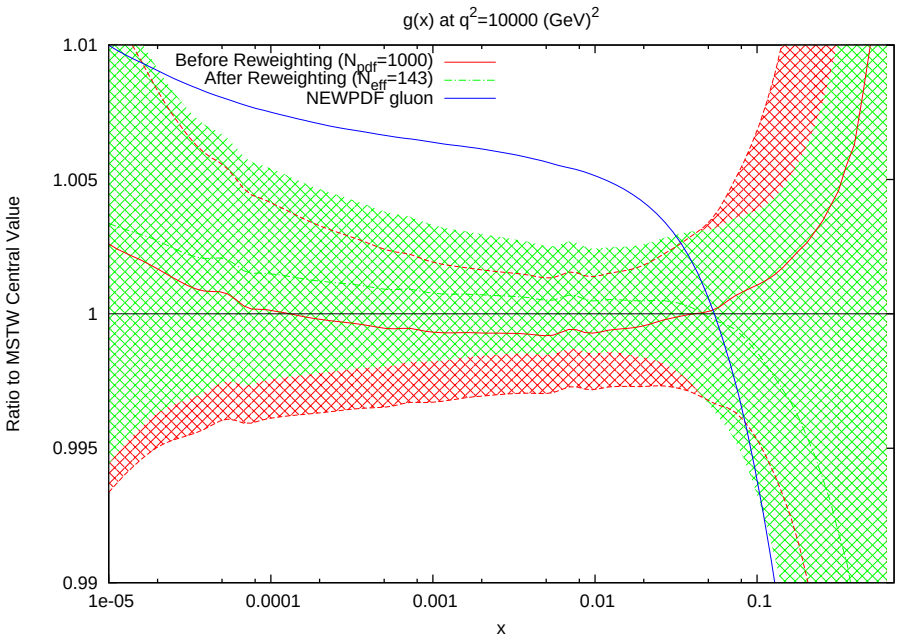


Figure 6.18.: Reweighted gluon using CMS inclusive data and the $\Delta\chi^2 = 1$ error treatment. The reweighting formula is a pure exponential.

6.5. Direct Inclusion of ATLAS 2.76 TeV + 7 TeV Data

The final new fit performed in this study is to include the ATLAS 2.76 TeV data in conjunction with the already present 7 TeV data. Whilst FastNLO tables for the 7 TeV data is available at the time of writing, this is not the case for the 2.76 TeV data. This presents the opportunity to interface APPLgrid, which did not exist at the time of the previous official fit, into the MSTW fitting code. This is an important development due to the wider range of cross sections beyond jet physics which can be used in APPLgrid. The process of writing an interface to the APPLgrid convolution code was a task of creating user defined α_s and PDF routines. With this achieved, it was possible to use the APPLgrid grids created in this thesis in a full MSTW fit.

Due to the fact that the MSTW code uses by default additive errors, the stringent cuts on the ATLAS ratio data discussed in Chapter 3 were applied to the data set

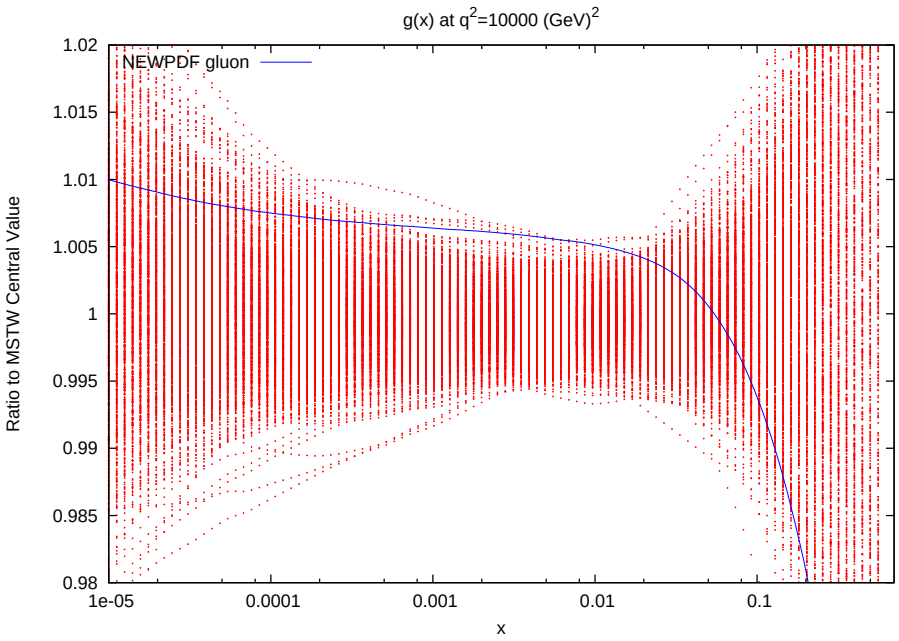


Figure 6.19.: Plot of the 1000 randomly distributed PDFs under the $\Delta\chi^2 = 1$ prescription

in the fit. APPLgrid grids were used for both of the ATLAS cross sections, and FastNLO was kept for all of the other jet cross sections, including the CMS inclusive jet data introduced in the previous section.

Again, α_s is allowed to initially go free, yielding an improvement of 20 fit points, and yielding a new value of $\alpha_s = 0.1187$. Most notable in this fit, shown as the second column in Table 6.2, is the very significant improvement in the ATLAS combined jet data. This improvement mostly goes away after holding α_s fixed at its MSTW2008 value. The total improvement in χ^2 in this case from the MSTW2008 NLO fit is 13 points, and so is better than the previous fit which only included the CMS and ATLAS 7 TeV data. The majority of the improvement is again caused by the CMS data which reduces by 14 points. The ATLAS combined data improves by 4 points, better than the 2 by the ATLAS 7 TeV in the previous fit, however there are 20 more points in the combined data set. This fit will be named here as MSTWATLAScomb, due to the additional inclusion of the combined ATLAS data. The fact that the CMS data improves more in this fit than the last demonstrates the good compatibility

Data Set	MSTW2008	MSTWATLAScomb α_s Free	MSTWATLAScomb α_s Fixed
BCDMS μp F_2	182/163	170/163	182/163
BCDMS μd F_2	190/151	189/151	190/151
NMC μp F_2	121/123	123/123	119/123
NMC μd F_2	102/123	103/123	101/123
NMC $\mu p/\mu d$	130/148	131/148	129/148
E665 μp F_2	57/53	53/53	54/53
E665 μd F_2	53/53	57/53	57/53
SLAC μp F_2	30/37	30/37	30/37
SLAC μd F_2	30/38	33/38	29/38
NMC/BCDMS/SLAC F_L	38/41	40/31	38/31
E866/NuSea pp DY	228/184	227/184	228/184
E866/NuSea pd/pp DY	14/15	13/15	14/15
NuTeV νN F_2	49/53	50/53	50/53
CHORUS νN F_2	26/42	26/42	26/42
NuTev νN xF_3	40/45	45/45	40/45
CHORUS νN xF_3	31/33	32/33	31/33
CCFFR $\nu N \to \mu\mu X$	66/86	67/86	65/86
NuTeV $\nu N \to \mu\mu X$	39/40	49/40	40/40
H1 MB 99 e^+p NC	9/8	9/8	9/8
H1 MB 97 e^+p NC	42/64	42/64	44/64
H1 low Q^2 96-97 e^+p NC	44/80	44/80	45/80
H1 high Q^2 98-99 e^-p NC	122/126	122/126	119/126
H1 high Q^2 99-00 e^+p NC	131/147	132/147	127/147
ZEUS SVX 95 e^+p NC	35/30	35/30	35/30
ZEUS 96-97 e^+p NC	86/144	86/144	85/144
ZEUS 98-99 e^-p NC	54/92	54/92	54/92
ZEUS 99-00 e^+p NC	63/90	63/90	62/90
H1 99-00 e^+p CC	29/28	29/38	29/28
ZEUS 99-00 e^+p CC	38/30	38/30	38/30
H1/ZEUS ep F_2^{charm}	107/83	105/83	109/83
H1 99-00 e^+p incl. jets	19/24	16/24	19/24
ZEUS 96-97 e^+p incl. jets	30/30	29/30	29/30
ZEUS 98-00 $e^\pm p$ incl. jets	17/30	16/30	17/30
D0 II $p\bar{p}$ incl. jets	114/110	116/110	116/110
CDF II $p\bar{p}$ incl. jets	56/76	63/76	58/76
CDF II $W \to l\nu$ asym.	29/22	29/22	29/22
D0 II $W \to l\nu$ asym.	25/10	28/10	25/10
D0 II Z rap.	19/28	18/28	19/28
CDF II Z rap.	49/29	49/29	50/29
ATLAS 2.76TeV and 7TeV incl. jets (R=0.4)	(149/114)	134/114	145/114
CMS 7 TeV incl. jets	(180/133)	161/133	166/133
Total	2872/2946	2852/2946	2859/2946

Table 6.2.: Table of χ^2 values for each data set included in the fits for the standard MSTW 2008 NLO fit and the new NLO fits with CMS and ATLAS combined 2.76 TeV and 7 TeV data. The ATLAS and CMS values are quoted for MSTW 2008 despite not being included in the fit. These are simply the χ^2 values obtained when the fit code is run using the standard set without minimisation.

between the two LHC scenarios. The ATLAS combined data, whilst only improving a small amount, is clearly having an additional affect on the global fit in the direction preferred by the CMS data.

Finally, the reweighting procedure can be once again tested against the direct inclusion of a new data set. This is achieved by reweighting the new MSTWCMS PDFs using the ATLAS ratio data. When comparing this to the change in the gluon by moving from MSTWCMS to MSTWATLAScomb, the results should agree if the two methods are consistent. The results for the gluon are shown in Fig 6.20. The agreement between the two methods is not as obvious as in the previous case with the inclusion of the CMS data, however the general trends are comparable, and both agree within their respective error bands. The MSTWATLAScomb fit is almost identical to MSTWCMS for most of the x range, with the only divergence coming at high x where the uncertainties are highest. This is testament to the dominance of the CMS data in both fits. The ATLAS ratio data has very little effect on its own when additionally added to the CMS fit. The top plot in Fig 6.20 shows that there is a small improvement in the error band of the PDFs when including the ATLAS ratio data, and so there is a benefit to including both data sets simultaneously.

6.6. Summary

For the first time, LHC data has been included in the MSTW PDF fit. The data sets included represent the highest precision inclusive jet cross sections from both ATLAS and CMS to date. Two fits were initially performed with the new data, including the CMS inclusive jet data and ATLAS 7 TeV inclusive jet data: one allowing all standard MSTW parameters to be free, and one with α_s fixed to the MSTW 2008 value. The entirely free set showed a significant reduction in global χ^2, although much of this was due to a shift in α_s which significantly improved some fixed target data. With α_s fixed the fit again improved, although to a lesser extent, with the majority of improvement coming from the new data sets. The improvement was dominated by the CMS data due to the previously noted issue of the large ATLAS systematic errors.

With the new central values and eigenvectors, the reweighting procedure was applied to study the change in the effect after inclusion of the ATLAS combined

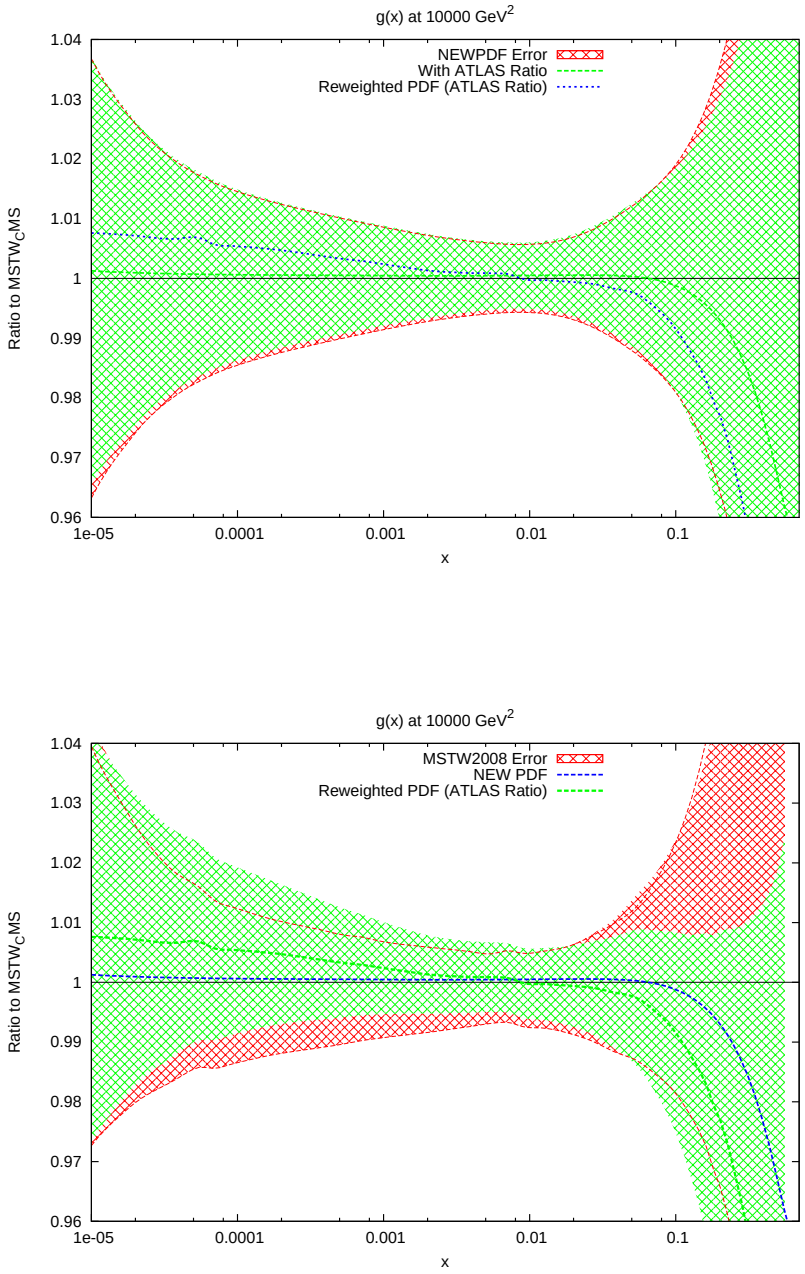


Figure 6.20.: Comparison of the gluon for the CMS fit, reweighted PDF (using ATLAS Ratio jets to reweight), and the new fit directly including the ATLAS Ratio & CMS data. The central values are the same on both plots; however the first plot shows the new PDF's error band in green, whilst the second shows the reweighed PDF's error band in green.

data set. This was shown to still have an effect on the gluon, with a similar but less pronounced shape than was seen when reweighting the MSTW2008 set with the same data. Dijet data was shown again to have a different effect on the PDFs to the corresponding inclusive data, which is further evidence for their value in a future global fit.

The ATLAS combined data, having shown an effect through reweighting, was then included in a second fit along with the CMS data. This further improved the global fit, with the fit to CMS showing a similar improvement to the first set, and an additional improvement from the ATLAS data itself. This demonstrated an agreement between the data sets, which had already been observed through the reweighting technique in Chapter 3.

Chapter 7.

Conclusions

This thesis has presented an overview of jet physics in relation to the structure of the proton. Firstly, deep inelastic scattering at HERA was studied in the context of parton shower Monte Carlo generators, with the aim of better understanding the higher order corrections due to soft QCD radiation. These corrections were shown to depend on the method of showering, with a pure LO matrix element with parton showering technique unable to properly describe inclusive jets at HERA at low momentum transfers. The methods which utilise hard emissions in the matrix element along with parton showers carefully constructed to avoid double counting were shown to describe the data excellently, including in the low momentum transfer region. Charged current inclusive jets were studied from the perspective of the valence quark PDFs and $s - \bar{s}$. This data, if sufficiently discriminating, could be used to separate the up and down quarks. A study of the effects on the MSTW2008 PDFs with NLO matrix elements in POWHEG showed that this data does indeed have a noticeable effect on the down quark, along with the strange distribution.

The first MSTW study using LHC inclusive jets has also been shown, with the latest data from ATLAS and CMS used to test the DGLAP evolution up to LHC energies. The first data set released was the ATLAS 7 TeV inclusive jet data, which showed an excellent fit to all PDFs. However, the large systematic uncertainties were shown to be the main reason for the good fit, with very little effect shown in reweighting the PDFs to the data. To reduce the systematic uncertainties, the 2.76 TeV and 7 TeV cross sections were simultaneously studied as a ratio. Many sources of uncertainty are correlated across the two data sets, and PDF effects were shown to be more apparent when fitting to them. A more realistic fit with a χ^2

of approximately 1 per point was seen for this data. A reweighting procedure was introduced which used the fit quality for a randomly produced PDF in eigenvector space as the weights. Using this, a shift in the gluon and reduction in error was seen to result. Finally, CMS inclusive jets at 7 TeV were studied. This data was released later than that from ATLAS, and hence had a higher luminosity. Again, the reweighting method showed a significant shift in the gluon for this set, and the change was comparable to the ATLAS ratio data. Sufficient motivation was found to create a new fit utilising these two data sets.

Dijet cross sections at hadron colliders were studied in Chapter 4. Previous studies had shown a lack of stability for the QCD predictions at the Tevatron, and so a full study was conducted of this data set and the new ATLAS data set to compare and contrast the calculations. The scale choice of $k * p_T^{av}$ with variable scale factor k was shown to break down in the high rapidity, high mass region due to the NLO correction becoming negative for low k. The use of $k * M_{JJ}$ and $k * M_{JJ}/0.7\cosh(y*)$ were shown to remove this instability, although the quality of fit was not always better for the D0 and ATLAS dijet data sets. A detailed study of the scale variations at different points in the phasespace showed that $M_{JJ}/0.7cosh(y*)$ was the most stable choice when considering all points in the ATLAS data set. By reweighting the data using differing scale choice calculations, the effect on the PDFs were shown to be significant but contraditory between the calculations. Therefore, whilst it would be useful to include dijet data alongside inclusive jet data in a PDF fit, these contradictions must be fully understood and brought under control.

Finally, various new PDF fits were performed. In Chapter 5, the MSTWCP and MSTWCPdeut fits were described and tested in relation to the ATLAS W/Z and W asymmetry data. The MSTW2008 PDFs had been shown to be inconsistent with this data, and the reason was an inability of the parametrisation of the valence distributions to obtain the correct $u_v - d_v$ distribution. A new parametrisation based on Chebyshev Polynomials was tested in the thesis and shown to improve the fit to the ATLAS W/Z data dramatically. The new parametrisation, however, was shown to have very little effect on the ATLAS jet data, due to the insensitivity of jet data to the $u_v - d_v$ distribution. Preliminary PDFs utilising this new parametrisation were produced using new data from HERA and the Tevatron, and again showed an excellent fit to the W/Z and asymmetry data from ATLAS.

New fits were then performed including the new jet data directly. The new fits including just the CMS and ATLAS 7 TeV data, named MSTWCMS due to the dominance of the CMS data, showed a change in the gluon consistent with that shown through reweighting of the MSTW2008 set, demonstrating directly for the first time the compatibility of the standard MSTW uncertainty and reweighting methods. The ATLAS combined 2.76 TeV and 7 TeV inclusive jet data was then additionally included in a full fit, MSTWATLAScomb. This fit showed a good agreement between the ATLAS and CMS data, with the CMS data improving a similar amount to the MSTWCMS fit, and the χ^2 for the ATLAS combined data simultaneously improving.

Appendix A.

NLOjet++ details

This section will go into the details of the NLOjet++ calculations used in this thesis. As discussed in Chapter 1, there are two currently available methods for using NLOjet++ with quick PDF and scale changes: FastNLO and APPLgrid. For most of the discussions in this thesis, APPLgrid has been the tool used. This is due to the existence of user code which may be used to produce predictions other than those produced by the code writers.

The two tools should, however, be in agreement when producing predictions for the same observables, since the underlying calculations use the same calculation tool. One difference that can produce disagreement is the exact details of the scale choice. For inclusive jets, FastNLO chooses the renormalisation and factorisation scales on an event by event basis using the p_T of the hardest jet in the event. APPLgrid, however, uses the highest p_T jet in the relevant rapidity bin.

The ratio between the two ATLAS inclusive jet predictions is shown at next-to leading order in Fig A.1 for three different PDFs. There is a disagreement between the two predictions, and the difference even depends on which PDFs are used. The disparity is at most 5%, mostly occurring at low p_T where scale effects are most dominant.

The obvious question to ask is whether the differences can affect PDF discrimination. Table A.1 shows the values for the ATLAS R=0.4 inclusive jet fit using both the FastNLO and APPLgrid predictions. The values are similar for each PDF, with the exception of ABM11 which gives the worst fit in both cases, but a much worse fit for FastNLO than APPLgrid. The ABM11 prediction also gives the biggest

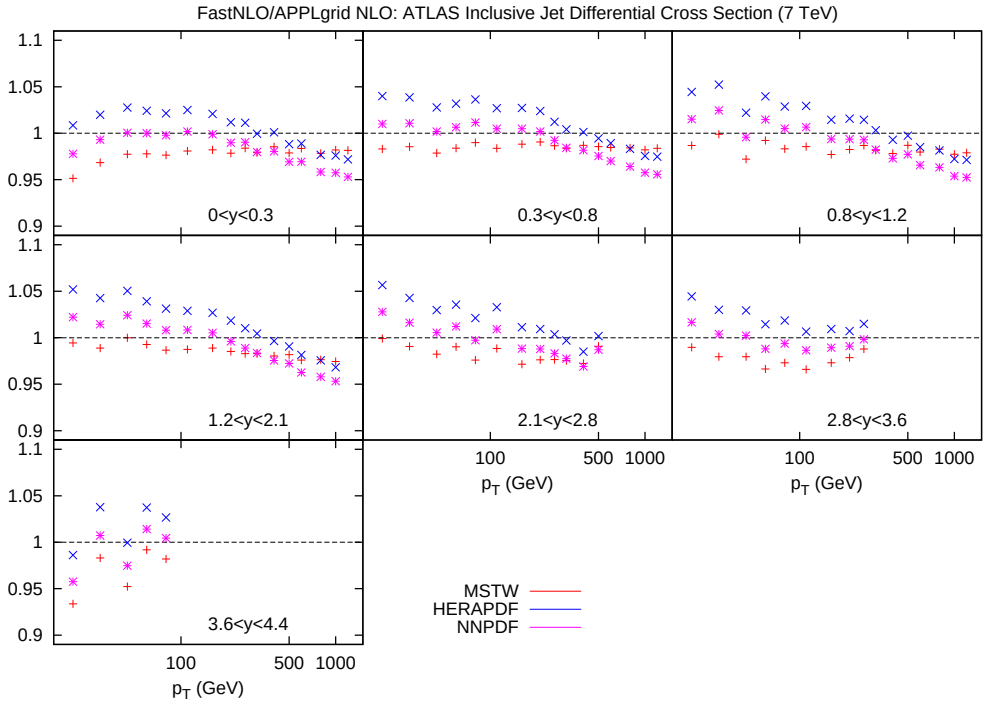


Figure A.1.: Ratio of FastNLO to APPLgrid predictions for the ATLAS R=0.4 inclusive jet cross section at NLO. The APPLgrid prediction used

difference in fit quality between the two calculation. Whilst the absolute change in the cross section is not more that other PDFs, most of the change occurs at low p_T, where the statistical errors are smallest, and so a change in the cross section here increases the χ^2 significantly. One must, then, be careful with which prediction is used for PDF comparisons. In this thesis, APPLgrid has been used throughout, and all quoted fit qualities are using these predictions.

It is necessary with any calculation to generate enough events to ensure that the theoretical prediction is not affected by statistical Monte Carlo errors. For the calculations performed using APPLgrid in this thesis, code was written to calculate the χ^2 fit to data as the statistics are improved. This information is shown in Fig A.2, which indicates the change in fit quality for ATLAS and D0 dijets as the number of events is increased. It is clear the plot evens out and becomes essentially flat after approximately 2 billion events. $9x10^8$ events were generated for the D0

	FastNLO	APPLgrid
MSTW 2008	0.599	0.584
CT10	0.595	0.587
NNPDF 22	0.582	0.595
HERAPDF 15	0.640	0.673
ABM11	1.48	0.989

Table A.1.: Comparison of fit quality for ATLAS inclusive jets (R=0.4) using APPLgrid and FastNLO predictions for each major PDF set.

calculation and $1.2x10^9$ events were generated for the ATLAS calculation. The ATLAS calculation takes longer to settle than D0, and so more events were generated for this data set. However, this plot demonstrates that it can be assumed that the χ^2 quoted for these calculations is a true reflection of the physics within the experimental statistical and systematic uncertainties.

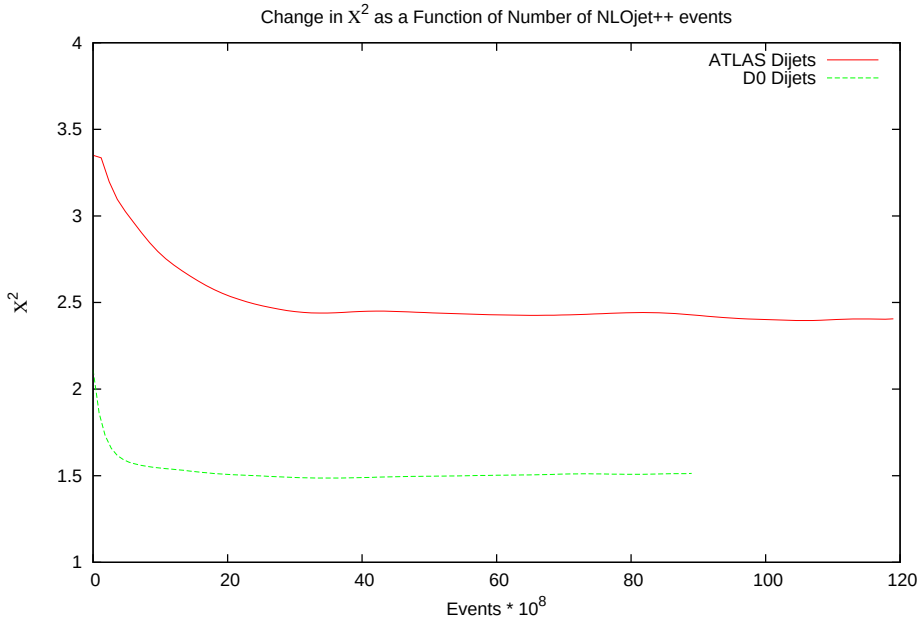


Figure A.2.: Plot of fit quality against events generated for ATLAS and D0 dijet calculations.

Bibliography

[1] J. Beringer et al. (Particle Data Group), Phys. Rev. D86, 010001 (2012)

[2] SLAC-MIT Collaboration, Phys. Rev. D5, 528 (1972)

[3] G. Altarelli and G. Parisi, Nuclear Physics B 126 2, 298-318 (1977)

[4] A. D. Martin W. J. Stirling, R. S. Thorne, G. Watt, Eur.Phys.J. C63 (2009) 189-285 [arXiv:0901.0002]

[5] NNPDF Collaboration, Nucl. Phys. B855:608-638 (2012) [arXiv:1108.1758]

[6] A. M. Cooper-Sarkar, PoS DIS2010:023 (2010) [arXiv:1006.4471]

[7] M. Guzzi et al., SMU-HEP-10-11 [arXiv:1101.0561]

[8] S. Alekhin, J. Bluemlein, S.-O. Moch, DESY 13-020 (2013) [arXiv:1302.1516]

[9] J. Pumplin et al., Phys. Rev. D 65 (2002) 014013 [arXiv:hep-ph/0101032]

[10] A. C. Benvenuti et al. [BCDMS Collaboration], Phys. Lett. B 223 (1989) 485.

[11] A. C. Benvenuti et al. [BCDMS Collaboration], Phys. Lett. B 237 (1990) 592.

[12] L. W. Whitlow, S. Rock, A. Bodek, E. M. Riordan and S. Dasu, Phys. Lett. B 250 (1990) 193.

[13] M. Arneodo et al. [New Muon Collaboration], Nucl. Phys. B 483 (1997) 3 [arXiv:hep-ph/9610231]

[14] M. Arneodo et al. [New Muon Collaboration], Nucl. Phys. B 487 (1997) 3 [arXiv:hep-ex/9611022]

[15] M. R. Adams et al. [E665 Collaboration], Phys. Rev. D 54 (1996) 3006.

[16] L. W. Whitlow, E. M. Riordan, S. Dasu, S. Rock and A. Bodek, Phys. Lett. B

282 (1992) 475.

[17] L. W. Whitlow, Ph.D. thesis, Stanford University, 1990, SLAC-0357.

[18] J. C. Webb, Ph.D. thesis, New Mexico State University, 2002, arXiv:hep-ex/0301031; Paul E. Reimer, private communication (for the radiative corrections).

[19] R. S. Towell et al. [FNAL E866/NuSea Collaboration], Phys. Rev. D 64 (2001) 052002 [arXiv:hep-ex/0103030]

[20] M. Tzanov et al. [NuTeV Collaboration], Phys. Rev. D 74 (2006) 012008 [arXiv:hep-ex/0509010]

[21] G. Onengut et al. [CHORUS Collaboration], Phys. Lett. B 632 (2006) 65.

[22] M. Goncharov et al. [NuTeV Collaboration], Phys. Rev. D 64 (2001) 112006 [arXiv:hep-ex/0102049]

[23] E. M. Lobodzinska [H1 Collaboration], [arXiv:hep-ph/0311180]

[24] C. Adloff et al. [H1 Collaboration], Eur. Phys. J. C 21 (2001) 33 [arXiv:hep-ex/0012053]

[25] C. Adloff et al. [H1 Collaboration], Eur. Phys. J. C 19 (2001) 269 [arXiv:hep-ex/0012052]

[26] C. Adloff et al. [H1 Collaboration], Eur. Phys. J. C 30 (2003) 1 [arXiv:hep-ex/0304003]

[27] J. Breitweg et al. [ZEUS Collaboration], Eur. Phys. J. C 7 (1999) 609 [arXiv:hep-ex/9809005]

[28] S. Chekanov et al. [ZEUS Collaboration], Eur. Phys. J. C 21 (2001) 443 [arXiv:hep-ex/0105090]

[29] S. Chekanov et al. [ZEUS Collaboration], Eur. Phys. J. C 28 (2003) 175 [arXiv:hep-ex/0208040]

[30] S. Chekanov et al. [ZEUS Collaboration], Phys. Rev. D 70 (2004) 052001 [arXiv:hep-ex/0401003]

[31] C. Adloff et al. [H1 Collaboration], Eur. Phys. J. C 30 (2003) 1 [arXiv:hep-

ex/0304003]

[32] S. Chekanov et al. [ZEUS Collaboration], Eur. Phys. J. C 32 (2003) 1 [arXiv:hep-ex/0307043]

[33] C. Adloff et al. [H1 Collaboration], Z. Phys. C 72 (1996) 593 [arXiv:hep-ex/9607012]

[34] C. Adloff et al. [H1 Collaboration], Phys. Lett. B 528 (2002) 199 [arXiv:hep-ex/0108039]

[35] A. Aktas et al. [H1 Collaboration], Eur. Phys. J. C 45 (2006) 23 [arXiv:hep-ex/0507081]

[36] A. Aktas et al. [H1 Collaboration], Eur. Phys. J. C 40 (2005) 349 [arXiv:hep-ex/0411046]

[37] J. Breitweg et al. [ZEUS Collaboration], Eur. Phys. J. C 12 (2000) 35 [arXiv:hep-ex/9908012]

[38] S. Chekanov et al. [ZEUS Collaboration], Phys. Rev. D 69 (2004) 012004 [arXiv:hep-ex/0308068]

[39] S. Chekanov et al. [ZEUS Collaboration], JHEP 0707 (2007) 074 [arXiv:0704.3562 [hep-ex]]

[40] A. Aktas et al. [H1 Collaboration], Phys. Lett. B 653 (2007) 134 [arXiv:0706.3722 [hep-ex]]

[41] S. Chekanov et al. [ZEUS Collaboration], Phys. Lett. B 547 (2002) 164 [arXiv:hep-ex/0208037]

[42] S. Chekanov et al. [ZEUS Collaboration], Nucl. Phys. B 765 (2007) 1 [arXiv:hep-ex/0608048]

[43] V. M. Abazov et al. [D0 Collaboration], Phys. Rev. Lett. 101 (2008) 062001 [arXiv:0802.2400 [hep-ex]]

[44] A. Abulencia et al. [CDF Collaboration], Phys. Rev. D 75 (2007) 092006 [Erratum- ibid. D 75 (2007) 119901] [arXiv:hep-ex/0701051]

[45] D. Acosta et al. [CDF Collaboration], Phys. Rev. D 71 (2005) 051104 [arXiv:hep-

ex/0501023]

[46] V. M. Abazov et al. [D0 Collaboration], Phys. Rev. D 77 (2008) 011106 [arXiv:0709.4254 [hep-ex]]

[47] V. M. Abazov et al. [D0 Collaboration], Phys. Rev. D 76 (2007) 012003 [arXiv:hep-ex/0702025]

[48] J. Han et al. [CDF Collaboration], May 2008, http://www-cdf.fnal.gov/physics/ewk/2008/dszdy/.

[49] G. Sterman and S. Weinberg, Phys. Rev. Lett. 39 1436 (1977)

[50] S. D. Ellis and D. Soper, Phys.Rev.D48:3160-3166 (1993) [arXiv:hep-ph/9305266]

[51] M. Cacciari, G. P. Salam and G. Soyez, JHEP04 063,(2008) [arXiv:0802.1189]

[52] ATLAS Collaboration, Phys.Rev. D86 (2012) 014022 [arXiv:1112.6297]

[53] CMS Collaboration, Phys. Rev. D 87 112002 (2013) [arXiv:1212.6660]

[54] M. Bahr et al. Eur. Phys. J. C58:639-707,(2008) [arXiv:0803.0883]

[55] T. Gleisberg et al., (2009) [arXiv:0811.4622]

[56] S. Hoeche, F. Krauss, S. Schumann and F. Siegert, JHEP05 053, (2009) [arXiv:0903.1219]

[57] S. Frixione, P. Nason, C. Oleari, JHEP11 070 (2007) [arXiv:0709.2092]

[58] Z. Nagy Phys. Rev. D68, 094002 (2003) [arXiv:0307268]

[59] Z. Nagy, Phys. Rev. Lett.88, 122003 (2002) [arXiv:0110315]

[60] T. Kluge, K. Rabbertz, M. Wobisch, in proceedings "14th International Workshop on Deep Inelastic Scattering" (2006) [arXiv:hep-ph/0609285]

[61] T. Carli et al., Eur. Phys. J. C 66 (2010) 503 [arXiv:0911.2985]

[62] S. Chekanov et al. [ZEUS Collaboration], Phys. Lett. B547:164-180 (2002) [arXiv:hep-ex/0208037]

[63] A. Buckley, J. Butterworth, D. Grellscheid, H. Hoeth, L. Lonnblad, J. Monk, H. Schulz and F. Siegert [arXiv:1003.0694]

[64] S. Chekanov et al. [ZEUS Collaboration], Phys. Rev. D78:032004, (2008) [arXiv:hep-ex/0208037]

[65] K. Mishra et al., Snowmass QCD Working Group (2013) [arXiv:1308.1430]

[66] CDF Collaboration, Phys.Rev.D75:09 (2006) [arXiv:hep-ex/0701051]

[67] D0 Collaboration, Phys.Rev.D85:05 (2006) [arXiv:1110.3771]

[68] D0 Collaboration, Phys. Lett. B693:531-538 (2010) [arXiv:1002.4594]

[69] ATLAS Collaboration, Eur. Phys. J. C 73:2509 (2013) [arXiv:1304.4739]

[70] W. T. Giele and S. Keller, Phys.Rev. D58 094023 (1998) [hep-ph/9803393]

[71] NNPDF Collaboration, R. D. Ball et al., Nucl. Phys. B849 112-143 (2011) [arXiv:1012.0836]

[72] NNPDF Collaboration, R. D. Ball et al., Nucl. Phys. B855 (2012) 608Ð638 [arXiv:1108.1758]

[73] G. Watt and R.S. Thorne, JHEP 1208:052 (2012) [arXiv:1205.4024]

[74] F. De Lorenzi, CERN-THESIS-2011-237 (2011)

[75] N. Kidonakis and J. F. Owens, Phys. Rev. D 63, 054019 (2001)

[76] A. Gehrmann-De Ridder et al. , Phys. Rev. Lett. 110, 162003 (2013)

[77] S. D. Ellis, Z. Kunszt and D. E. Soper, Phys. Rev. Lett. 69, 1496-1499 (1992)

[78] S. D. Ellis,et al., Phys. Rev. Lett. 69, 1496 (1992)

[79] J. Huston - https://indico.cern.ch/contributionDisplay.py?contribId=3&confId=226756

[80] R. S. Thorne and G. Watt, JHEP 1108:100 (2011) [arXiv:1106.5789]

[81] ATLAS Collaboration, Phys. Rev. D85:072004 (2012) [arXiv:1109.5141]

[82] A. D. Martin, A. J. Th. M. Mathijssen, W. J. Stirling, R. S. Thorne, B. J. A. Watt, G. Watt, Eur. Phys. J. C 73:2318 (2013) [arXiv:1211.1215]

[83] R. S. Thorne, A. D. Martin, W. J. Stirling, G. Watt, PoS DIS2010:052 (2010) [arXiv:1006.2753]

[84] R. S. Thorne, Phys.Rev. D86:074017 (2012) [arXiv:1201.6180]

[85] H1 Collaboration, ZEUS Collaboration, JHEP 1001:109 (2010) [arXiv:0911.0884]

[86] H1 Collaboration, ZEUS Collaboration, DESY-12-172 (2010) [arXiv:1211.1182]

[87] ZEUS Collaboration, Eur.Phys.J.C62:625-658, (2009) [arXiv:0901.2385]

[88] H1 Collaboration, Eur.Phys.J.C71:1579 (2011) [arXiv:1012.4355]

[89] H1 Collaboration, Phys.Lett.B665:139-146 (2008) [arXiv:0805.2809]

[90] ZEUS Collaboration, Phys.Lett.B682:8-22 (2009) [arXiv:0904.1092]

[91] D0 Collaboration, Phys.Rev.Lett. 101:211801 (2008) [arXiv:0807.3367]

[92] CDF Collaboration, Phys.Rev.Lett. 102:181801 (2009) [arXiv:0901.2169]

[93] CDF Collaboration, Phys.Lett.B692:232-239 (2010) [arXiv:0908.3914]

[94] D. Britzger, K. Rabbertz, F. Stober and M. Wobisch, Proceedings of the XX International Workshop on Deep Inelastic Scattering, University of Bonn, 26-30th March 2012 [arXiv:1208.3641]